

(21) (A1) **2,186,833**

(22) 1996/09/30

(43) 1998/03/31

(72) Gijzen, Mark, CA

(71) HER MAJESTY IN RIGHT OF CANADA AS REPRESENTED BY  
THE MINISTER OF AGRICULTURE, CA

(51) Int.Cl.<sup>6</sup> C12N 15/33, A01H 5/00, A01H 5/10

(54) **ADN GENOMIQUE ET PEROXYDASE SPECIFIQUE POUR  
TEGUMENTS**

(54) **SEED COAT DNA REGULATORY REGION AND PEROXIDASE**

(57) Caractérisation et présentation d'une nouvelle séquence génomique spécifique pour le tégument. Les régions régulatrices voisines de l'ADN ont également été caractérisées. Le peroxydase de tégument est traduit sous forme de protéine précurseur de 38 kDa, à 352 acides aminés, renfermant une séquence-signal de 26 acides aminés; elle donne, par clivage, une protéine de 35 kDa. Les plantes renfermant un allèle Ep dominant accumulent de grandes quantités de peroxydase dans les cellules sabliers du subépiderme. Les génotypes epep homozygotes récessifs n'accumulent pas de peroxydase dans ces cellules et leur part dans l'activité totale de la peroxydase du tégument se trouve sensiblement réduite. Les sondes dérivées de l'ADNc ou de l'ADN génomique peuvent servir à déceler les polymorphismes qui distinguent les génotypes EpEp et epep. La coségrégation des polymorphismes dans une population F<sub>2</sub> provenant d'un croisement de plantes EpEp et epep montre que le locus Ep code la protéine peroxydase. Une comparaison des allèles Ep et ep révèle qu'il manque 87 bp dans le gène récessif pour le codon initial de traduction. L'expression hétérologue ainsi que les vecteurs et les hôtes utilisés pour l'expression de la peroxydase du tégument sont également présentés. La région régulatrice de l'ADN spécifique pour la semence peut servir à contrôler l'expression i) de certains gènes, comme ceux codant la résistance aux herbicides, ii) de protéines virales du tégument, protégeant contre l'infection, iii) de protéines à intérêt commercial (p. ex. en pharmacie), iv) de protéines modifiant la valeur nutritive, le goût ou le conditionnement des semences; enfin, elle peut servir à v) éliminer biologiquement des insectes ou des agents pathogènes (p. ex. B. thuringiensis).

(57) A novel seed coat specific peroxidase genomic sequence is characterized and presented. Adjacent DNA regulatory regions have also been characterized. The seed coat peroxidase is translated as a 352 amino acid precursor protein of 38 kDa comprising a 26 amino acid signal sequence which when cleaved results in a 35 kDa protein. Plants containing a dominant Ep allele accumulate large amounts of peroxidase in the hourglass cells of the subepidermis. Homozygous recessive epep genotypes do not accumulate peroxidase in the hourglass cells and are much reduced in total seed coat peroxidase activity. Probes derived from the cDNA, or genomic DNA can be used to detect polymorphisms that distinguished EpEp and epep genotypes. Cosegregation of the polymorphisms in an F<sub>2</sub> population from a cross of EpEp and epep plants shows that the Ep locus encodes the seed coat peroxidase protein. Comparison of Ep and ep alleles indicates that the recessive gene lacks 87 bp of sequence encompassing the translation start codon. The heterologous expression, as well as vectors and hosts to be used for the expression of the seed coat peroxidase, are also disclosed. The seed-specific DNA regulatory region maybe used to control expression of genes of interest such as i) genes encoding herbicide resistance, or ii) biological control of insects or pathogens (e.g. B. thuringiensis), or iii) viral coat proteins to protect against viral infections, or iv) proteins of commercial interest (e.g. pharmaceutical), and v) proteins that alter the nutritive value, taste, or processing of seeds.



**ABSTRACT OF THE DISCLOSURE**

A novel seed coat specific peroxidase genomic sequence is characterized and presented. Adjacent DNA regulatory regions have also been characterized. The seed coat peroxidase is translated as a 352 amino acid precursor protein of 38 kDa comprising a 26 amino acid signal sequence which when cleaved results in a 35 kDa protein. Plants containing a dominant *Ep* allele accumulate large amounts of peroxidase in the hourglass cells of the subepidermis. Homozygous recessive *epep* genotypes do not accumulate peroxidase in the hourglass cells and are much reduced in total seed coat peroxidase activity. Probes derived from the cDNA, or genomic DNA can be used to detect polymorphisms that distinguished *EpEp* and *epep* genotypes. Cosegregation of the polymorphisms in an  $F_2$  population from a cross of *EpEp* and *epep* plants shows that the *Ep* locus encodes the seed coat peroxidase protein. Comparison of *Ep* and *ep* alleles indicates that the recessive gene lacks 87 bp of sequence encompassing the translation start codon. The heterologous expression, as well as vectors and hosts to be used for the expression of the seed coat peroxidase, are also disclosed. The seed-specific DNA regulatory region may be used to control expression of genes of interest such as i) genes encoding herbicide resistance, or ii) biological control of insects or pathogens (e.g. *B. thuringiensis*), or iii) viral coat proteins to protect against viral infections, or iv) proteins of commercial interest (e.g. pharmaceutical), and v) proteins that alter the nutritive value, taste, or processing of seeds.

5

**SEED COAT DNA REGULATORY REGION AND PEROXIDASE**

The present invention relates to a novel DNA molecule comprising a plant seed coat specific DNA regulatory region and a novel structural gene encoding a peroxidase. The seed-coat specific DNA regulatory region may also be used to control the expression of other genes of interest within the seed coat.

10

**BACKGROUND OF THE INVENTION**

Full citations for references appear at the end of the Examples section.

15

Peroxidases are enzymes catalyzing oxidative reactions that use  $H_2O_2$  as an electron acceptor. These enzymes are widespread and occur ubiquitously in plants as isozymes that may be distinguished by their isoelectric points. Plant peroxidases contribute to the structural integrity of cell walls by functioning in lignin biosynthesis and suberization, and by forming covalent cross-linkages between

20

extensin, cellulose, pectin and other cell wall constituents (Campa, 1991). Peroxidases are also associated with plant defence responses and resistance to pathogens (Bowles, 1990; Moerschbacher 1992). Soybeans contain 3 anionic isozymes of peroxidase with a minimum  $M_r$  of 37 kDa (Sessa and Anderson, 1981).

25

Recently one peroxidase isozyme, localised within the seed coat of soybean, has been characterized with a  $M_r$  of 37 kDa (Gillikin and Graham, 1991).

In an analysis of soybean seeds, Buttery and Buzzell (1968) showed that the amount of peroxidase activity present in seed coats may vary substantially among different cultivars. The presence of a single dominant gene *Ep* causes a high seed

different cultivars. The presence of a single dominant gene *Ep* causes a high seed coat peroxidase phenotype (Buzzell and Buttery, 1969). Homozygous recessive *epep* plants are ~100-fold lower in seed coat peroxidase activity. This results from a reduction in the amount of peroxidase enzyme present, primarily in the hourglass cells of the subepidermis (Gijzen *et al.*, 1993). In plants carrying the *Ep* gene, peroxidase is heavily concentrated in the hourglass cells (osteosclereids). These cells form a highly differentiated cell layer with thick, elongated secondary walls and large intercellular spaces (Baker *et al.*, 1987). Hourglass cells develop between the epidermal macrosclereids and the underlying articulated parenchyma, and are a prominent feature of seed coat anatomy at full maturity. The cytoplasm exudes from the hourglass cells upon imbibition with water and a distinct peroxidase isozyme constitutes five to 10% of the total soluble protein in *EpEp* seed coats. It is not known why the hourglass cells accumulate large amounts of peroxidase, but the sheer abundance and relative purity of the enzyme in soybean seed coats is significant because peroxidases are versatile enzymes with many commercial and industrial applications. Studies of soybean seed coat peroxidase have shown this enzyme to have useful catalytic properties and a high degree of thermal stability even at extremes of pH (McEldoon *et al.*, 1995). These properties result in the preferred use of soybean peroxidase, over that of horseradish peroxidase, in diagnostic assays as an enzyme label for antigens, antibodies, oligonucleotide probes, and within staining techniques. Johnson *et al* report on the use of soybean peroxidase for the deinking of printed waste paper (U.S. 5,270,770; December 6, 1994) and for the biocatalytic oxidation of primary alcohols (U.S. 5,391,488; February 13, 1996). Soybean peroxidase has also been used as a replacement for

chlorine in the pulp and paper industry, or as formaldehyde replacement (Freiberg, 1995).

5       An anionic soybean peroxidase from seed coats has been purified (Gillikin and Graham, 1991). This protein has a pI of 4.1 and  $M_r$  of 37 kDa. A method for the bulk extraction of peroxidase from seed hulls of soybean using a freeze thaw technique has also been reported (U.S. 5,491,085, February 13, 1996, Pokara and Johnson).

10       Lagrimini et al (1987) disclose the cloning of a ubiquitous anionic peroxidase in tobacco encoding a protein of  $M_r$  of 36 kDa. This peroxidase has also been over expressed in transgenic tobacco plants (Lagrimini et al 1990) and Maliyakal discloses the expression of this gene in cotton (WO 95/08914).

15       Huangpu et al (1995) reported the partial cloning of a soybean anionic seed coat peroxidase. The 1031 bp sequence contained an open reading frame of 849 bp encoding a 283 amino acid protein with a  $M_r$  of 30,577. The  $M_r$  of this peroxidase is 7 kDa less than what one would expect for a soybean seed coat peroxidase as reported by Gillikin and Graham (1991) and possibly represents another peroxidase  
20       isozyme within the seed coat.

The upstream promoter sequences for two poplar peroxidases have been described by Osakabe et al (1995). A number of characteristic regulatory sites were identified from comparison of these sequences to existing promoter elements.

Additionally, a cryptic promoter with apparent specificity for seed coat tissues was isolated from tobacco by a promoter trapping strategy (Fobert et al. 1994). The upstream regulatory sequences associated with the Ep gene in soybean are distinct from these and other previously characterized promoters. The soybean Ep promoter  
5 drives high-level expression in a cell and tissue specific manner. The peroxidase protein encoded by the Ep gene accumulates in the seed coat tissues, especially in the hour glass cells of the subepidermis. Minimal expression of the gene is detected in root tissues.

10

One problem arising from the desired use of soybean seed coat peroxidase is that there is variability between soybean varieties regarding peroxidase production (Buttery and Buzzell, 1986; Freiberg, 1995). Due to the commercial interest in the use of soybean seed coat peroxidase new methods of producing this enzyme are  
15 required. Therefore, the gene responsible for the expression of the 37 kDa isozyme in soybean seed coat was isolated and characterized.

Furthermore, novel regulatory regions obtained from the genomic DNA of soybean seed coat peroxidase have been isolated and characterized and are useful  
20 in directing the expression of genes of interest in seed coat tissues.

#### SUMMARY OF THE INVENTION

The present invention relates to a DNA molecule that encodes a soybean seed coat peroxidase and associated DNA regulatory regions.

This invention also embraces isolated DNA molecules having the nucleotide sequence of either SEQ ID NO:1 (the cDNA encoding soybean seed coat peroxidase) or SEQ ID No:2 (the genomic sequence).

5        This invention also provides for a chimeric DNA molecule comprising a seed coat-specific regulatory region having nucleotides 1-191 of SEQ ID NO:2 and a gene of interest under control of this DNA regulatory region. Also included within this invention are chimeric DNA molecules comprising genomic DNA sequences exemplified by nucleotides 412-1041, 1234-2263 or 2430-2691 of SEQ ID NO:2.

10

The present invention also provides for vectors which comprise DNA molecules encoding soybean seed coat peroxidase. Such a construct may include the DNA regulatory region from SEQ ID NO:2 in conjunction with the seed coat peroxidase gene, or the seed coat peroxidase gene under the control of any suitable  
15        constitutive or inducible promoter of interest.

This invention is also directed towards vectors which comprise a gene of interest placed under the control of a DNA regulatory element derived from the genomic sequence encoding soybean seed coat peroxidase. Such a regulatory  
20        element includes nucleotides 1-191 of SEQ ID NO:2. Elements comprising nucleotides 412-1041, 1234-2263 or 2430-2691 of SEQ ID NO:2 may also be used.

This invention also embraces prokaryotic and eukaryotic cells comprising the vectors identified above. Such cells may include bacterial, insect, mammalian, and plant cell cultures.

5           This invention also provides for transgenic plants comprising the seed coat peroxidase gene under control of constitutive or inducible promoters. Furthermore, this invention also relates to transgenic plants comprising the DNA regulatory regions of nucleotides 1-191 of SEQ ID NO:2 controlling a gene of interest, or comprising genes of interest in functional association with genomic DNA sequences  
10       exemplified by nucleotides 412-1041, 1234-2263 or 2430-2691 of SEQ ID NO:2.

This invention is also directed to a method for the production of soybean seed coat peroxidase in a host cell comprising:

- 15           i) transforming the host cell with a vector comprising an oligonucleotide sequence that encodes soybean seed coat peroxidase; and  
            ii) culturing the host cell under conditions to allow expression of the soybean seed coat peroxidase.

20           This invention also provides for a process for producing a heterologous gene of interest within seed coats of a transformed plant, comprising propagating a plant transformed with a vector comprising a gene of interest under the control of nucleotides 1-191 of SEQ ID NO:2



Although the present invention is exemplified by a soybean seed coat peroxidase and adjacent DNA regulatory regions, in practice any gene of interest can be placed downstream from the DNA regulatory region for seed coat specific expression.

**BRIEF DESCRIPTION OF THE DRAWINGS**

These and other features of the invention will become more apparent from the following description in which reference is made to the appended drawings wherein:

5

**Figure 1** is the cDNA and deduced amino acid sequence of soybean seed coat peroxidase. Nucleotides are numbered by assigning +1 to the first base of the ATG start codon; amino acids are numbered by assigning +1 to the N-terminal Gln residue after cleavage of the putative signal sequence. The N-terminal signal sequence, the region of the active site, and the heme-binding domain are underlined. The numerals I, II and III placed directly above single nucleotide gaps in the sequence indicate the three intron splice positions. The target site and direction of five different PCR primers are shown with dotted lines above the nucleotide sequence. An asterisk (\*) marks the translation stop codon.

10

15

**Figure 2** is the genomic DNA sequence of the Soybean seed coat peroxidase.

**Figure 3** is a comparison of soybean seed coat peroxidase with other closely related plant peroxidases. The GenBank accession numbers are provided next to the name of the plant from which the peroxidase was isolated. The accession number for the soybean sequence is L78163. (A) A comparison of the nucleic acid sequences; (B) A comparison of the amino acid sequences.

20

Figure 4 is a restriction fragment length polymorphisms between *EpEp* and *epep* genotypes using the seed coat peroxidase cDNA as probe. Genomic DNA of soybean lines OX312 (*epep*) and OX347 (*EpEp*) was digested with restriction enzyme, separated by electrophoresis in a 0.5% agarose gel, transferred to nylon, and hybridized with <sup>32</sup>P-labelled cDNA encoding the seed coat peroxidase. The size of the hybridizing fragments was estimated by comparison to standards and is indicated on the right.

Figure 5 exhibits the structure of the *Ep* Locus. A 17 kb fragment including the *Ep* locus is illustrated schematically. A 3.3 kb portion of the gene is enlarged and exons and introns are represented by shaded and open boxes, respectively. The final enlargement of the 5' region shows the location and DNA sequence around the 87 bp deletion occurring in the *ep* allele of soybean line OX312. Nucleotides are numbered by assigning +1 to the first base of the ATG start codon.

Figure 6 displays PCR analysis of *EpEp* and *epep* genotypes using primers derived from the seed coat peroxidase cDNA. Genomic DNA from soybean lines OX312 (*epep*) and OX347 (*EpEp*) was used as template for PCR analysis with four different primer sets. Amplification products were separated by electrophoresis through a 0.8% agarose gel and visualized under UV light after staining with ethidium bromide. Genotype and primer combinations are indicated at the top of the figure. The size in base pairs of the amplified DNA fragments are indicated on the right.

Figure 7 exhibits PCR analysis of an F<sub>2</sub> population from a cross of *EpEp* and *epep* genotypes. Genomic DNA was used as template for PCR analysis of the parents (P) and 30 F<sub>2</sub> individuals. The cross was derived from the soybean lines OX312 (*epep*) and OX347 (*EpEp*). Plants were self pollinated and seeds were collected and scored for seed coat peroxidase activity. The symbols (-) and (+) indicate low and high seed coat peroxidase activity, respectively. Primers prx9+ and prx10- were used in the amplification reactions. Products were separated by electrophoresis through a 0.8% agarose gel and visualized under UV light after staining with ethidium bromide. The migration of molecular markers and their corresponding size in kb is also shown (lanes M).

Figure 8 displays PCR analysis of six different soybean cultivars with primers derived from the seed coat peroxidase cDNA sequence. Genomic DNA was used as template for PCR analysis of three *EpEp* cultivars and three *epep* cultivars. Primers used in the amplification reactions and the size of the DNA product is indicated on the left. Products were separated by electrophoresis through a 0.8% agarose gel and visualized under UV light after staining with ethidium bromide.

- (A) Forward and reverse primers are downstream from deletion  
(B) Forward primer anneals to site within deletion  
(C) Primers span deletion

**DESCRIPTION OF PREFERRED EMBODIMENT**

The present invention is directed to a novel oligonucleotide sequence encoding a seed coat peroxidase and associated DNA regulatory regions.

5       According to the present invention DNA sequences that are "substantially homologous" includes sequences that are identified under conditions of high stringency. "High stringency" refers to Southern hybridization conditions employing washes at 65°C with 0.1 x SSC, 0.5 % SDS.

10       By "DNA regulatory region" it is meant any region within a genomic sequence that has the property of controlling the expression of a DNA sequence that is operably linked with the regulatory region. Such regulatory regions may include promoter or enhancer regions, and other regulatory elements recognized by one of skill in the art. A segment of the DNA regulatory region is exemplified in this  
15       invention, however, as is understood by one of skill in the art, this region may be used as a probe to identify surrounding regions involved in the regulation of adjacent DNA, and such surrounding regions are also included within the scope of this invention.

20       In the context of this disclosure, the term "promoter" or "promoter region" refers to a sequence of DNA, usually upstream (5') to the coding sequence of a structural gene, which controls the expression of the coding region by providing the recognition for RNA polymerase and/or other factors required for transcription to start at the correct site.

There are generally two types of promoters, inducible and constitutive. An "inducible promoter" is a promoter that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. In the absence of an inducer the DNA sequences or genes will not be transcribed.

- 5 Typically the protein factor, that binds specifically to an inducible promoter to activate transcription, is present in an inactive form which is then directly or indirectly converted to the active form by the inducer. The inducer can be a chemical agent such as a protein, metabolite, growth regulator, herbicide or phenolic compound or a physiological stress imposed directly by heat, cold, salt, or toxic  
10 elements or indirectly through the action of a pathogen or disease agent such as a virus. A plant cell containing an inducible promoter may be exposed to an inducer by externally applying the inducer to the cell or plant such as by spraying, watering, heating or similar methods.

- 15 By "constitutive promoter" it is meant a promoter that directs the expression of a gene throughout the various parts of a plant and continuously throughout plant development. Examples of known constitutive promoters include those associated with the CaMV 35S transcript and *Agrobacterium* Ti plasmid nopaline synthase gene.

20

The chimeric gene constructs of the present invention can further comprise a 3' untranslated region. A 3' untranslated region refers to that portion of a gene comprising a DNA segment that contains a polyadenylation signal and any other regulatory signals capable of effecting mRNA processing or gene expression. The

polyadenylation signal is usually characterized by effecting the addition of polyadenylic acid tracks to the 3' end of the mRNA precursor. Polyadenylation signals are commonly recognized by the presence of homology to the canonical form 5' AATAAA-3' although variations are not uncommon.

5

Examples of suitable 3' regions are the 3' transcribed non-translated regions containing a polyadenylation signal of *Agrobacterium* tumour inducing (Ti) plasmid genes, such as the nopaline synthase (*Nos* gene) and plant genes such as the soybean storage protein genes and the small subunit of the ribulose-1, 5-bisphosphate carboxylase (ssRUBISCO) gene. The 3' untranslated region from the structural gene of the present construct can therefore be used to construct chimeric genes for expression in plants.

The chimeric gene construct of the present invention can also include further enhancers, either translation or transcription enhancers, as may be required. These enhancer regions are well known to persons skilled in the art, and can include the ATG initiation codon and adjacent sequences. The initiation codon must be in phase with the reading frame of the coding sequence to ensure translation of the entire sequence. The translation control signals and initiation codons can be from a variety of origins, both natural and synthetic. Translational initiation regions may be provided from the source of the transcriptional initiation region, or from the structural gene. The sequence can also be derived from the promoter selected to express the gene, and can be specifically modified so as to increase translation of the mRNA.

To aid in identification of transformed plant cells, the constructs of this invention may be further manipulated to include plant selectable markers. Useful selectable markers include enzymes which provide for resistance to an antibiotic such as gentamycin, hygromycin, kanamycin, and the like. Similarly, enzymes providing for production of a compound identifiable by colour change such as *GUS* (β-glucuronidase), or luminescence, such as luciferase are useful.

Also considered part of this invention are transgenic plants containing the chimeric gene construct of the present invention. Methods of regenerating whole plants from plant cells are known in the art, and the method of obtaining transformed and regenerated plants is not critical to this invention. In general, transformed plant cells are cultured in an appropriate medium, which may contain selective agents such as antibiotics, where selectable markers are used to facilitate identification of transformed plant cells. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be used to establish repetitive generations, either from seeds or using vegetative propagation techniques.

The constructs of the present invention can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, micro-injection, electroporation, etc. For reviews of such techniques see for example Weissbach and Weissbach (1988) and Geierson and Corey (1988). The



present invention further includes a suitable vector comprising the chimeric gene construct.

Buttery and Buzzell (1968) showed that the amount of peroxidase activity  
5 present in seed coats may vary substantially among different cultivars. The presence  
of a single dominant gene *Ep* causes a high seed coat peroxidase phenotype (Buzzell  
and Buttery, 1969). Homozygous recessive *epep* plants are ~100-fold lower in seed  
coat peroxidase activity. This results from a reduction in the amount of peroxidase  
enzyme present, primarily in the hourglass cells of the subepidermis (Gijzen *et al.*,  
10 1993). In plants carrying the *Ep* gene, peroxidase is heavily concentrated in the  
hourglass cells (osteosclereids). These cells form a highly differentiated cell layer  
with thick, elongated secondary walls and large intercellular spaces (Baker *et al.*,  
1987).

15 Screening a seed coat cDNA library prepared from *EpEp* plants with a  
degenerate primer derived from the active site domain of plant peroxidase resulted  
in a high frequency of positive clones. Many of these clones encode identical  
cDNA molecules and indicate that the corresponding mRNA is an abundant  
transcript in developing seed coat tissues. The sequence of the cDNA is shown in  
20 Figure 1.

Previous studies on soybean seed coat peroxidase indicated that this enzyme  
is heavily glycosylated and that carbohydrate contributes 1% of the mass of the  
apo-enzyme (Gray *et al.*, 1996). The seven potential glycosylation sites identified

from the amino acid sequence of the seed coat peroxidase (Figure 1) would accommodate the five or six N-linked glycosylation sites proposed by Gray *et al.* (1996). The heme-binding domain encompasses residues Asp161 to Phe171 and the acid-base catalysis region from Gly33 to Cys44. The two regions are highly conserved among plant peroxidases and are centred around functional histidine residues, His169 and His40. There are eight conserved cysteine residues in the mature protein that provide for four di-sulfide bridges found in other plant peroxidases and predicted from the crystal structure of peanut peroxidase (Welinder, 1992; Schuller *et al.*, 1996). Other conserved areas include residues Cys91 to Ala105 and Val119 to Leu127 that occur in or around helix D. The most divergent aspects of the seed coat peroxidase protein sequence are the carboxy- and amino-terminal regions. These sequences probably provide special targeting signals for the proper processing and delivery of the peptide chain. It is possible the carboxy-terminal extension of the seed coat peroxidase is removed at maturity, as has been shown for certain barley and horseradish peroxidases (Welinder, 1992).

The molecular mass of the enzyme has been determined by denaturing gel electrophoresis to be 37 kDa (Sessa and Anderson, 1981; Gillikin and Graham, 1991) or 43 kDa (Gijzen *et al.*, 1993). Analysis by mass spectrometry indicated a mass of 40,622 Da for the apo-enzyme and 33,250 Da after deglycosylation (Gray *et al.*, 1996). These values are in good agreement with the mass of 35,377 Da calculated from the predicted amino acid sequence for the mature apo-protein prior to glycosylation and other modifications. Huangpu *et al.* (1995) reported an anionic seed coat peroxidase having a  $M_r$  of 30,577 Da and characterized a partial cDNA

encoding this protein. This 1031 bp cDNA contained an open reading frame of 849 bp encoding a 283 amino acid protein. There are several differences between this reported sequence and the sequence of this invention that are manifest at the amino acid level (see Figure 3 for sequence comparison). The enzyme encoded by the gene reported by Huangpu et al is different from that of this invention as the peroxidase of this invention has a  $M_r$  of 35,377 Da.

Genomic DNA blots probed with the seed coat peroxidase cDNA produced two or three hybridizing fragments of varying intensity with most restriction enzyme digestions, despite that several peroxidase isozymes are present in soybean. The results indicate that this seed coat peroxidase is present as a single gene that does not share sufficient homology with most other peroxidase genes to anneal under conditions of high stringency.

The genomic DNA sequence (Figure 2) comprises four exons spanning bp 191-411 (exon I), 1042 -1233 (exon 2), 2264-2429 (exon 3) and 2692-3174 (exon 4) and three introns comprising 412-1041 (intron 1), 1234-2263 (intron 2) and 2430-2691 (intron 3). Features of the upstream regulatory region of the genomic DNA include a TATA box centred on bp 147; a cap signal 32 bp down stream centred on bp 179. Also noted within the genomic sequence are three polyadenylation signals centred on bp 3180, 3258, 3323 and a polyadenylation site at bp 3359.

This promoter is considered seed coat specific since the peroxidase protein encoded by the Ep gene accumulates in the seed coat tissues, especially in the

hourglass cells of the subepidermis, and is not expressed in other tissues, aside from a marginal expression of peroxidase in the root tissues. The DNA regulatory regions of the genomic sequence of Figure 2 are used to control the expression of the adjacent peroxidase gene in seed coat tissue. Such regulatory regions include

5 nucleotides 1-191. Other regions of interest include nucleotides 412-1041, 1234-2263 and/or 2430-2691 of SEQ ID NO:2. Therefore other proteins of interest may be expressed in seed coat tissues by placing a gene capable of expressing the protein of interest under the control of the DNA regulatory elements of this invention. Genes of interest include but are not restricted to herbicide resistant genes, genes

10 encoding viral coat proteins, or genes encoding proteins conferring biological control of pest or pathogens such as an insecticidal protein for example *B. thuringiensis* toxin. Other genes include those capable of the production of proteins that alter the taste of the seed and/or that affect the nutritive value of the soybean.

15 A modified DNA regulatory sequence may be obtained by introducing changes into the natural sequence. Such modifications can be done through techniques known to one of skill in the art such as site-directed mutagenesis, reducing the length of the regulatory region using endonucleases or exonucleases, increasing the length through the insertion of linkers or other sequences of interest.

20 Reducing the size of DNA regulatory region may be achieved by removing 3' or 5' regions of the regulatory region of the natural sequence by using a endonuclease such as BAL 31 (Sambrook et al 1989). However, any such DNA regulatory region must still function as a seed coat specific DNA regulatory region.

It may be readily determined if such modified DNA regulatory elements are capable of acting in a seed coat specific manner transforming plant cells with such regulatory elements controlling the expression of a suitable marker gene, culturing these plants and determining the expression of the marker gene within the seed coat as outlined above. One may also analyze the efficacy of DNA regulatory elements by introducing constructs comprising a DNA regulatory element of interest operably linked with an appropriate marker into seed coat tissues by using particle bombardment directed to seed coat tissue and determining the degree of expression of the regulatory region (reference).

10

Two tandemly arranged genes encoding anionic peroxidase expressed in stems of *Populus kitakamiensis*, *prxA3a* and *prxA4a* have been cloned and characterized (Osakabe et al, 1995). Both of these genomic sequences contained four exons and three introns and encoded proteins of 347 and 343 amino acids, respectively. The two genes encode distinct isozymes with deduced  $M_r$ s of 33.9 and 34.6 kDa. Furthermore, a 532 bp promoter derived from the peroxidase gene of *Armoracia rusticana* has also been reported (Toyobo KK, JP 4,126,088, April 27, 1992). However, a search using GenBank revealed no substantial similarity between the promoter region, or introns 1, 2 and 3 of this invention and those within the literature.

20

Digestion of the genomic DNA with *Bam*HI or *Sac*I revealed restriction fragment length polymorphisms that distinguished *EpEp* and *epep* genotypes. Although the *Xba*I digestion did not produce a readily detectable polymorphism, the

size of the hybridizing fragment in both genotypes was ~14 kb. Thus, a 0.3 kb size difference is outside of the resolving power of the separation for fragments this large. Sequence analysis of *EpEp* and *epep* genotypes indicates that the mutant *ep* allele is missing 87 bp of sequence at the 5' end of the structural gene. This would  
5 account for the drastically reduced amounts of peroxidase enzyme present in seed coats of *epep* plants since the deletion includes the translation start codon and the entire N-terminal signal sequence. However, the 87 bp deletion cannot account for the differences observed in the RFLP analysis since the missing fragment does not include a *Bam*HI site and is much smaller than the 0.3 kb polymorphism detected  
10 in the *Sac*I digestion. Thus, other genetic rearrangements must occur in the vicinity of the *ep* locus that lead to these polymorphisms.

The results shown here indicate that the mutation causing low seed coat peroxidase activity occurs in the structural gene encoding the enzyme. This mutation  
15 is an 87 bp deletion in the 5' region of the gene encompassing the translation start site. Several different low peroxidase cultivars share a similar mutation in the same area, suggesting that the recessive *ep* alleles have a common origin or that the region is prone to spontaneous deletions or rearrangements.

20 Due to the industrial interest in soybean seed coat peroxidase, alternate sources for the production of this enzyme are needed. The DNA of this invention, encoding the seed coat soybean peroxidase under the control of a suitable promoter and expressed within a host of interest, can be used for the preparation of recombinant soybean seed coat peroxidase enzyme.

Soybean seed coat peroxidase has been characterized as a lignin-type peroxidase that has industrially significant properties ie: high activity and stability under acidic conditions; exhibits wide substrate specificity; equivalent catalytic properties to that of *Phanerochaete chrysosporium* lignin peroxidase (the currently preferred enzyme used for treatment of industrial waste waters (Wick 1995) but is at least 150-fold more stable; more stable than horseradish peroxidase which is also used in industrial effluent treatments and medical diagnostic kits (McEldoon *et al.*, 1995). These properties are useful within industrial applications for the degradation of natural aromatic polymers including lignin and coal (McEldoon *et al.*, 1995), and the preferred use of soybean peroxidase, over that of horseradish peroxidase, in medical diagnostic tests as an enzyme label for antigens, antibodies, oligonucleotide probes, and within staining techniques (Wick 1995). Soybean peroxidase is also used in the deinking of printed waste paper (Johnson *et al.*, U.S. 5,270,770; December 6, 1994) and for the biocatalytic oxidation of primary alcohols (Johnson *et al.*, U.S. 5,391,488; February 13, 1996). Soybean peroxidase has also been used as a replacement for chlorine in the pulp and paper industry, in order to remove chlorine, phenolic or aromatic amine containing pollutants from industrial waste waters (Wick 1995), or as formaldehyde replacement (Freiberg, 1995) for use in adhesives, abrasives, and protective coatings (e.g. varnish and resins, Wick 1995).

20

Furthermore, the seed coat peroxidase gene may be expressed in an organ or tissue specific manner within a plant. For example, the quality and strength of cotton fiber can be improved through the over-expression of cotton or horseradish

peroxidase placed under the control of a fibre-specific promoter (Maliyakal, WO 95/08914; April 6, 1995).

Similarly, seed-specific DNA regulatory regions of this invention may be  
5 used to control expression of genes of interest such as:

- i) genes encoding herbicide resistance, or
  - ii) biological control of insects or pathogens (e.g, *B. thuringiensis*), or
  - iii) viral coat proteins to protect against viral infections, or
  - iv) proteins of commercial interest (e.g. pharmaceutical), and
  - 10 v) proteins that alter the nutritive value, taste, or processing of seeds
- within the seed coat of plants.

While this invention is described in detail with particular reference to preferred embodiments thereof, said embodiments are offered to illustrate but not  
15 to limit the invention.

## EXAMPLES

### *Plant material*

20

All soybean (*Glycine max* [L.] Merr) cultivars and breeding lines were from the collection at Agriculture Canada, Harrow, Ontario.



*Seed Coat cDNA library Construction and Screening*

High seed coat peroxidase (*EpEp*) soybean cultivar Harosoy 63 plants were grown in field plots outdoors. Pods were harvested 35 days after flowering and seeds in the mid-to-late developmental stage were excised. The average fresh mass was 250 mg per seed. Seed coats were dissected and immediately frozen in liquid nitrogen. The frozen tissue was lyophilized and total RNA extracted in 100 mM Tris-HCl pH 9.0, 20 mM EDTA, 4% (w/v) sarkosyl, 200 mM NaCl, and 16 mM DTT, and precipitated with LiCl using the standard phenol/chloroform method described by Wang and Vodkin (1994). The poly(A)<sup>+</sup> RNA was purified on oligo(dT) cellulose columns prior to cDNA synthesis, size selection, ligation into the  $\lambda$  ZAP Express vector, and packaging according to instructions (Stratagene). A degenerate oligonucleotide with the 5' to 3' sequence of TT(C/T)CA(C/T)GA(C/T)TG(C/T)TT(C/T)GT was 5' end labelled to high specific activity and used as a probe to isolate peroxidase cDNA clones (Sambrook *et al.*, 1989). Duplicate plaque lifts were made to nylon filters (Amersham), UV fixed, and prehybridized at 36 °C for 3 h in 6 x SSC, 20 mM Na<sub>2</sub>HPO<sub>4</sub> (pH6.8), 5 x Denhardt's, 0.4 % SDS, and 500  $\mu$ g/mL salmon sperm DNA. Hybridization was in the same buffer, without Denhardt's, at 36 °C for 16 h. Filters were washed quickly with several changes of 6 x SSC and 0.1 % SDS, first at room temperature and finally at 40°C, prior to autoradiography for 16 h at -70°C with an intensifying screen.

*Genomic DNA Isolation, Library Construction, and DNA Blot Analysis*

Soybean genomic DNA was isolated from leaves of greenhouse grown plants or from etiolated seedlings grown in vermiculite. Plant tissue was frozen in liquid nitrogen and lyophilized before extraction and purification of DNA according to the method of Dellaporta *et al.* (1983). Restriction enzyme digestion of 30  $\mu$ g DNA, separation on 0.5 % agarose gels and blotting to nylon membranes followed standard protocols (Sambrook *et al.*, 1989). For construction of the genomic library, DNA purified from Harosoy 63 leaf tissue was partially digested with *Bam*HI and ligated into the  $\lambda$  FIX II vector (Stratagene). Gigapack XL packaging extract (Stratagene) was used to select for inserts of 9 to 22 kb. After library amplification, duplicate plaque lifts were hybridized to cDNA probe.

Blots or filter lifts were prehybridized for 2 h at 65°C in 6 x SSC, 5 x Denhardt's, 0.5 % SDS, and 100  $\mu$ g/mL salmon sperm DNA. Radiolabelled cDNA probe (20 to 50 ng) was prepared using the Ready-to-Go labelling kit (Pharmacia) and  $^{32}$ P-dCTP (Amersham). Unincorporated  $^{32}$ P-dCTP was removed by spin column chromatography before adding radiolabelled cDNA to the hybridization buffer (identical to prehybridization buffer without Denhardt's). Hybridization was for 20 h at 65°C. Membranes were washed twice for 15 min at room temperature with 2 x SSC, 0.5 % SDS, followed by two 30 min washes at 65°C with 0.1 x SSC, 0.5 % SDS. Autoradiography was for 20 h at -70°C using an intensifying screen and X-OMAT film (Kodak).

*DNA Sequencing*

Sequencing of DNA was performed using dye-labelled terminators and Taq-FS DNA polymerase (Perkin-Elmer). The PCR protocol consisted of 25 cycles of  
5 a 30 sec melt at 96°C, 15 sec annealing at 50°C, and 4 min extension at 60°C. Samples were analyzed on an Applied Biosystems 373A Stretch automated DNA sequencer.

*Polymerase Chain Reaction*

10

PCR amplifications contained 1 ng template DNA, 5 pmol each primer, 1.5 mM MgCl<sub>2</sub>, 0.15 mM deoxynucleotide triphosphates mix, 10 mM Tris-HCl, 50 mM KCl, pH 8.3, and 1 unit of Taq polymerase (Gibco BRL) in a total volume of 25 µL. Reactions were performed in a Perkin-Elmer 480 thermal cycler. After an initial  
15 2 min denaturation at 94°C, there were 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 52°C, and 2 min extension at 72°C. A final 7 min extension at 72°C completed the program. The following primers were used for PCR analysis of genomic DNA:

prx2+	CTTCCAAATATCAACTCAAT
20 prx6-	TAAAGTTGGAAAAGAAAGTA
prx9+	ATGCATGCAGGTTTTTCAGT
prx10-	TTGCTCGCTTTCTATTGTAT
prx12+	TCTTCGATGCTTCTTTCACC
prx29+	CATAAACAATACGTACGTGAT

*Seed Coat Peroxidase Assays*

The F<sub>3</sub> seed was measured for peroxidase activity to score the phenotype of the F<sub>2</sub> population because the seed testa is derived from maternal tissue. The seeds  
5 were briefly soaked in water and the seed coat was dissected from the embryo and placed in a vial. Ten drops (~500 µL) of 0.5% guaiacol was added and the sample was left to stand for 10 min before adding one drop (~50 µL) of 0.1% H<sub>2</sub>O<sub>2</sub>. An immediate change in colour of the solution, from clear to red, indicates a positive result and high seed coat peroxidase activity.

10

**Example 1:** *The Seed Coat Peroxidase cDNA and genomic DNA sequences*

To isolate the seed coat peroxidase transcript, a cDNA library was constructed from developing seed coat tissue of the *EpEp* cultivar Harosóy 63. The  
15 primary library contained 10<sup>6</sup> recombinant plaque forming units and was amplified prior to screening. A degenerate 17-mer oligonucleotide corresponding to the conserved active site domain of plant peroxidases was used to probe the library. In screening 10,000 plaque forming units, 12 positive clones were identified. The cDNA insert size of the clones ranged from 0.5 to 2.5 kb, but six clones shared a  
20 common insert size of 1.3 kb. These six clones (*soyprx03*, *soyprx05*, *soyprx06*, *soyprx11*, *soyprx12*, and *soyprx14*) were chosen for further characterization since the 1.3 kb insert size matched the expected peroxidase transcript size. Sequence analysis of the six clones showed that they contained identical cDNA transcripts encoding

a peroxidase and that each resulted from an independent cloning event since the junction between the cloning vector and the transcript was different in all cases.

Since it was not clear that the entire 5' end of the cDNA transcript was  
5 complete in any of the cDNA clones isolated, the structural gene corresponding to  
the seed coat peroxidase was isolated from a Harosoy 63 genomic library. A partial  
*Bam*HI digest of genomic DNA was used to construct the library and more than 10<sup>6</sup>  
plaque forming units were screened using the cDNA probe. A positive clone, G25-2-  
1-1-1, containing a 17 kb insert was identified and a 3.3 kb region encoding the  
10 peroxidase was sequenced (Figure 2).

The genomic sequence matched the cDNA sequence except for three introns  
encoded within the gene. The genomic sequence also revealed two additional  
translation start codons, beginning one bp and 10 bp upstream from the 5' end of  
15 the longest cDNA transcript isolated. Figure 1 shows the deduced cDNA sequence.  
The open reading frame of 1056 bp encodes a 352 amino acid protein of 38,106 Da.  
A heme-binding domain, a peroxidase active site signature sequence, and seven  
potential N-glycosylation sites were identified from the deduced amino acid  
sequence. The first 26 amino acid residues conform to a membrane spanning  
20 domain. Cleavage of this putative signal sequence releases a mature protein of 326  
residues with a mass of 35,377 Da and an estimated pI of 4.4.

Relevant features of the genomic fragment (Figure 2) include four exons at  
bp 192-411 (exon 1), 1042 -1233 (exon 2), 2263-2429 (exon 3) and 2692-3174

(exon 4) and three introns at bp 412-1041 (intron 1), 1234-2263 (intron 2) and 2430-2691 (intron 3). The 191 bp regulatory region of the genomic DNA include a TATA box centred on bp 147 and a cap signal 32 bp down stream centred at bp 179. Also noted within the genomic sequence are three polyadenylation signals  
5 centred on bp 3180, 3258, 3323 and a polyadenylation site at bp 3359.

Figure 3 illustrates the relationship between the soybean seed coat peroxidase and other selected plant peroxidases. The soybean sequence is most closely related to four peroxidase cDNAs isolated from alfalfa, (see Figure 3) sharing from 65 to  
10 67% identity at the amino acid level with the alfalfa proteins (X90693, X90694, X90692, el-Turk et al 1996; L36156, Abrahams et al 1994). When compared with other plant peroxidases, soybean seed coat peroxidase exhibits from 60 to 65% identity with poplar (D30653 and D30652, Osakabe et al 1994)) and flax (L0554, Omann and Tyson 1995); 50 to 60% identity with horseradish (M37156, Fujiyama  
15 et al. 1988), tobacco (D11396, Osakabe et al 1993), and cucumber (M91373, Rasmussen et al. 1992); and 49% identity with barley (L36093, Scott-Craig et al. 1994), wheat (X85228, Baga et al 1995) and tobacco (L02124, Diaz-De-Leon et al 1993) peroxidases.

20 **Example 2:**        *DNA Blot Analysis Using the Seed Coat Peroxidase cDNA Probe Reveals Restriction Fragment Length Polymorphisms Between EpEp and epep Genotypes*

Genomic DNA blots of OX347 (*EpEp*) and OX312 (*epep*) plants were hybridized with  $^{32}\text{P}$ -labelled cDNA to estimate the copy number of the seed coat peroxidase gene and to determine if this locus is polymorphic between the two genotypes. Figure 4 shows the hybridization patterns after digestion with *Bam*HI, *Xba*I, and *Sac*I. Restriction fragment length polymorphisms are clearly visible in the *Bam*HI and *Sac*I digestions. The *Bam*HI digestion produced a strongly hybridizing 17 kb fragment and a faint 3.4 kb fragment in the *EpEp* genotype. The 3.4 kb *Bam*HI fragment is visible in the *epep* genotype but the 17 kb fragment has been replaced by a signal at >20 kb. The *Sac*I digestion resulted in detection of three fragments in *EpEp* and *epep* plants. At least two fragments were expected here since the cDNA sequence has a *Sac*I site within the open reading frame. However, the smallest and most strongly hybridizing of these fragments is 5.2 kb in *EpEp* plants and 4.9 kb in *epep* plants. Digestion with *Xba*I produced hybridizing fragments of ~14 kb and 7.8 kb for both genotypes, with the larger fragment showing a stronger signal.

**Example 3:**      *A Deletion Mutation Occurs in the Recessive ep Locus*

The structural gene encoding the seed coat peroxidase is schematically illustrated in Figure 5. The 17 kb *Bam*HI fragment encompassing the gene includes 191 bp of sequence upstream from the translation start codon, three introns of 631 bp, 1030 bp, and 263 bp, and 13 kb of sequence downstream from the polyadenylation site. The arrangement of four exons and three introns and the

placement of introns within the sequence is similar to that described for other plant peroxidases (Simon, 1992; Osakabe *et al.* 1995).

Primers were designed from the DNA sequence to compare *EpEp* and *epep* genotypes by PCR analysis. Figure 6 shows PCR amplification products from four different primer combinations using OX312 (*epep*) and OX347 (*EpEp*) genomic DNA as template. The primer annealing site for prx29+ begins 182 bp upstream from the ATG start codon; the remaining primer sites are shown in Figure 1. Amplification with primers prx2+ and prx6-, and with prx12+ and prx10- produced the expected products of 1.9 kb and 860 bp, respectively, regardless of the *Ep/ep* genotype of the template DNA. However, PCR amplification with primers prx9+ and prx10-, and with prx29+ and prx10- generated the expected products only when template DNA was from plants carrying the dominant *Ep* allele. When template DNA was from an *epep* genotype, no product was detected using primers prx9+ and prx10- and a smaller product was amplified with primers prx29+ and prx10-. The products resulting from amplification of OX312 or OX347 template DNA with primers prx29+ and prx10- were directly sequenced and compared. The polymorphism is due to an 87 bp deletion occurring within this DNA fragment in OX312 plants, as shown in Figure 5. This deletion begins nine bp upstream from the translation start codon and includes 78 bp of sequence at the 5' end of the open reading frame, including the prx9+ primer annealing site.

To test whether this deletion mutation cosegregates with the seed coat peroxidase phenotype, genomic DNA from an F<sub>2</sub> population segregating at the *Ep*



locus was amplified using primers prx9+ and prx10- and F<sub>3</sub> seed was tested for seed coat peroxidase activity. Figure 7 shows the results from this analysis. Of the 30 F<sub>2</sub> individuals tested, all 23 that were high in seed coat peroxidase activity produced the expected 860 bp PCR amplification product. The remaining seven F<sub>2</sub>'s with low  
5 seed coat peroxidase activity produced no detectable PCR amplification products.

Finally, to determine if the OX312(*epep*) and OX347(*EpEp*) breeding lines are representative of soybean cultivars that differ in seed coat peroxidase activity, several cultivars were tested by PCR analysis using primer combinations targeted  
10 to the *Ep* locus. Figure 8 shows results from this analysis of six different soybean cultivars, three each of the homozygous dominant *EpEp* and recessive *epep* genotypes. As observed with OX312 and OX347, amplification products of the expected size were produced with primers prx12+ and prx10- regardless of the genotype, whereas *epep* genotypes yielded no product with primers prx9+ and  
15 prx10- or a smaller fragment with primers prx29+ and prx10-.

All scientific publications and patent documents are incorporated herein by reference.

20 The present invention has been described with regard to preferred embodiments. However, it will be obvious to persons skilled in the art that a number of variations and modifications can be made without departing from the scope of the invention as described in the following claims.

## References

- Abrahams, S.L., Hayes, C.M., and Watson, J.M. (1994) Organ-specific expression of three peroxidase-encoding cDNAs from lucerne (*Medicago sativa*). GenBank Accession # L36156.
- 5 Baga, M. Chibbar, R.N., and Kartha, K.K., (1995) Molecular cloning and expression analysis of peroxidase genes from wheat. *Plant Molec. Biol.* **29**, 647-662
- (1994) Baker, D.M., Minor, H.C., and Cumbie, B.G. (1987) Scanning electron microscopy examination of soybean testa development. *Can. J. Bot.*  
10 **65**, 2420-2424.
- Bowles, D.J. (1990) Defense-related proteins in higher plants. *Annu. Rev. Biochem.* **59**, 873-907.
- Buttery, B.R., and Buzzell, R.I. (1968) Peroxidase activity in the seeds of soybean varieties. *Crop Sci.* **8**, 722-725.
- 15 Buzzell, R.I., and Buttery, B.R. (1969) Inheritance of peroxidase activity in soybean seed coats. *Crop Sci.* **9**, 387-388.
- Campa, A. (1991) Biological roles of plant peroxidases: known and potential function. In *Peroxidases in Chemistry and Biology*, Volume II (J. Everse, K.E. Everse and M.B. Grisham, eds). Boca Raton, FL: CRC Press, pp. 25-  
20 50.
- Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983) A plant DNA miniprep. Version II. *Plant Mol. Biol. Rep.* **1**, 19-21.

- Diaz-De-Leon, f., Klotz, K.L., andf Lagrimini, M. (1993) Nucleotide Sequence of the Tobacco (*Nicotiana tabacum*) anionic peroxidase gene. *Plant Physiol.* 101, 1117-1118.
- 5 el-Turk, J., Asemota, O., Leymarie, J., Sallaud, C., Mesnage, S., Breda, C., Buffard, D., Kondorosi, A., and Esnault, R. (1996) Nucleotide sequence of four pathogen-induced alfalfa peroxide-encoding cDNAs. *Gene* 170, 213-216.
- Freiberg B., (1995) Inadiana Crop: Keeping Its Members Up with the Changing Times. *Seed Crops Indust.* March, 4-9
- 10 Fujiyama, K., Takemura, H., Shibayama, S., Kobayashi, K., Choi, J.-K., Shinmyo, A., Takano, M., Yamada, Y., and Okada, H. (1988) Structure fo the Horseradish Perosidase isozyme c genes. *Eur. J. Biochem.* 173, 681-687.
- 15 Geierson and Corey (1988), *Plant Molecular Biology*, 2d Ed. Gijzen, M., van Huystee, R., and Buzzell, R.I. (1993) Soybean seed coat peroxidase. A comparison of high-activity and low-activity genotypes. *Plant Physiol.* 103, 1061-1066.
- 20 Gillikin, J.W., and Graham, J.S. (1991) Purification and developmental analysis of the major anionic peroxidase from the seed coat of soybean. *Plant Physiol.* 96, 214-220.
- Gray, J.S.S., Yang, B.Y., Hull, S.R., Venzke, D.P., and Montgomery, R. (1996) The glycans of soybean peroxidase. *Glycobiology* 6, 23-32.

- Lagrimini, M.L., Bradford., and Rothstein S, (1990)** Peroxidase-Induced Wilting in Transgenic Tobacco. *Plant Cell* 2, 7-18.
- McEldoon, J.P., Pokora A.R., and Dordick, J.S. (1995)** Lignin peroxidase-type activity of soybean peroxidase. *Enzyme Microb. Technol.* 17, 359-365.
- 5 **Moerschbacher, B.M. (1992)** Plant peroxidases: involvement in response to pathogens. In *Plant Peroxidases 1980-1990: Topics and Detailed Literature on Molecular, Biochemical, and Physiological Aspects*, (C. Penel, T. Gaspar and H. Greppin, eds). Geneva: University of Geneva, pp. 91-115.
- Omman, F., and Tyson, H., (1995)** cDNA sequence of a peroxidase from flax  
10 (*Linum usitatissimum*), GenBank Accession # L07554.
- Osakabe, K., Koyama, H., Kawai, S., Katayama, Y., and Morohoshi, N. (1993)** Nucleotide sequence for the genomic DNA encoding the anionic peroxidase gene from *Nicotiana tabacum*. GenBank Accession # D11396.
- Osakabe, K., Koyama, H., Kawai, S., Katayama, Y., and Morohoshi, N. (1994)**  
15 Molecular cloning and nucleotide sequences of two novel cDNA that encode anionic peroxidases of *Populus kitakamiensis*. GenBank Accession # D30652.
- Osakabe, K., Koyama, H., Kawai, S., Katayama, Y., and Morohoshi, N. (1995)** Molecular cloning of two tandemly arranged peroxidase genes from *Populus kitakamiensis* and their differential regulation in the stem. *Plant Mol. Biol.*  
20 28, 677-689.
- Rasmussen, J.B., Smith, J.A., Williams, S., Burkhart, W., Ward, E.R., Somerville, S.C., Ryals, J., and Hammerschmidt, R. (1992)** Cloning and

Systemic Expression of an acidic peroxidase associated with systemic acquired resistance to disease in cucumber. GenBank Accession # M91373.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor, NY: Cold Spring Harbor  
5 Laboratory Press.

Schuller, D.J., Ban, N., van Huystee, R.B., McPherson, A., and Poulos, T.L.  
(1996) The crystal structure of peanut peroxidase. *Structure* 4, 311-321.

Scott-Craig, J.S., Kerby, K.B., Stein, B.D., and Sommerville, S.C. (1994)  
Expression of an extracellular peroxidase that is induced in barley (*Hordeum  
10 vulgare*) by the powdery mildew pathogen (*Erysiphe graminis f. sp. hordei*).  
GenBank Accession # L36093.

Sessa, D.J., and Anderson, R.L. (1981) Soybean peroxidases: Purification and  
some properties. *J. Agric. Food Chem.* 29, 960-965.

Simon, P. (1992) Molecular cloning of plant peroxidases. In *Plant Peroxidases  
15 1980-1990: Topics and Detailed Literature on Molecular, Biochemical, and  
Physiological Aspects* (C. Penel, T. Gaspar and H. Greppin, eds) Geneva:  
University of Geneva, pp. 47-58.

Wang, C.S., and Vodkin, L.O. (1994) Extraction of RNA from tissues containing  
high levels of procyanidins that bind RNA. *Plant Mol. Biol. Rep.* 12, 132-  
20 145.

Weissbach and Weissbach, (1988) *Methods for Plant Molecular Biology*, Academy  
Press, New York VIII, pp. 421-463

Wick, C.B. (1995) Enzymol International Shows Promise of Novel Peroxidase,  
*Chem. Eng. News*, pp. 1

**Welinder, K.G.** (1992) Plant peroxidase structure-function relationships. In *Plant Peroxidases 1980-1990: Topics and Detailed Literature on Molecular, Biochemical, and Physiological Aspects* (C. Penel, T. Gaspar and H. Greppin, eds) Geneva: University of Geneva, pp. 1-24.

## SEQUENCE LISTING

- (1) GENERAL INFORMATION:
- (i) APPLICANT:
- 5 (A) NAME: Mark Gijzen  
 (B) STREET: 848 Princess Avenue  
 (C) CITY: London  
 (D) STATE: Ontario  
 (E) COUNTRY: Canada  
 10 (F) POSTAL CODE (ZIP): N5W 3M4
- (ii) TITLE OF INVENTION: Seed Coat DNA Regulatory Region and Peroxidase
- (iii) NUMBER OF SEQUENCES: 2
- (iv) COMPUTER READABLE FORM:
- 15 (A) MEDIUM TYPE: Floppy disk  
 (B) COMPUTER: IBM PC compatible  
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
- 20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1244 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- 25 (ii) MOLECULE TYPE: cDNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 30 (B) LOCATION:1..1056  
 (ix) FEATURE:  
 (A) NAME/KEY: sig\_peptide  
 (B) LOCATION:1..77
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
- 35 ATG GGT TCC ATG CGT CTA TTA GTA GTG GCA TTG TTG TGT GCA TTT GCT  
 Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys Ala Phe Ala
- 1 5 10 15 48

- 38 -

	ATG CAT GCA GGT TTT TCA GTC TCT TAT GCT CAG CTT ACT CCT ACG TTC	96
	Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr Pro Thr Phe	
	20 25 30	
	TAC AGA GAA ACA TGT CCA AAT CTG TTC CCT ATT GTG TTT GGA GTA ATC	144
5	Tyr Arg Glu Thr Cys Pro Asn Leu Phe Pro Ile Val Phe Gly Val Ile	
	35 40 45	
	TTC GAT GCT TCT TTC ACC GAT CCC CGA ATC GGG GCC AGT CTC ATG AGG	192
	Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser Leu Met Arg	
	50 55 60	
10	CTT CAT TTT CAT GAT TGC TTT GTT CAA GGT TGT GAT GGA TCA GTT TTG	240
	Leu His Phe His Asp Cys Phe Val Gln Gly Cys Asp Gly Ser Val Leu	
	65 70 75 80	
	CTG AAC AAC ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA CTT CCA AAT	288
	Leu Asn Asn Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu Pro Asn	
15	85 90 95	
	ATC AAC TCA ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC AAG ACA GCG	336
	Ile Asn Ser Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys Thr Ala	
	100 105 110	
	GTG GAA AAT AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT ATT CTT GCT	384
20	Val Glu Asn Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile Leu Ala	
	115 120 125	
	ATT GCA GCT GAA ATA GCT TCT GTT CTG GGA GGA GGT CCA GGA TGG CCA	432
	Ile Ala Ala Glu Ile Ala Ser Val Leu Gly Gly Gly Pro Gly Trp Pro	
	130 135 140	
25	GTT CCA TTA GGA AGA AGG GAC AGC TTA ACA GCA AAC CGA ACC CTT GCA	480
	Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala Asn Arg Thr Leu Ala	
	145 150 155 160	
	AAT CAA AAC CTT CCA GCA CCT TTC TTC AAC CTC ACT CAA CTT AAA GCT	528
	Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu Thr Gln Leu Lys Ala	
30	165 170 175	
	TCC TTT GCT GTT CAA GGT CTC AAC ACC CTT GAT TTA GTT ACA CTC TCA	576
	Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp Leu Val Thr Leu Ser	
	180 185 190	
	GGT GGT CAT ACG TTT GGA AGA GCT CGG TGC AGT ACA TTC ATA AAC CGA	624
35	Gly Gly His Thr Phe Gly Arg Ala Arg Cys Ser Thr Phe Ile Asn Arg	
	195 200 205	
	TTA TAC AAC TTC AGC AAC ACT GGA AAC CCT GAT CCA ACT CTG AAC ACA	672



2186833

- 39 -

Leu Tyr Asn Phe Ser Asn Thr Gly Asn Pro Asp Pro Thr Leu Asn Thr  
 210 215 220  
 ACA TAC TTA GAA GTA TTG CGT GCA AGA TGC CCC CAG AAT GCA ACT GGG 720  
 Thr Tyr Leu Glu Val Leu Arg Ala Arg Cys Pro Gln Asn Ala Thr Gly  
 5 225 230 235 240  
 GAT AAC CTC ACC AAT TTG GAC CTG AGC ACA CCT GAT CAA TTT GAC AAC 768  
 Asp Asn Leu Thr Asn Leu Asp Leu Ser Thr Pro Asp Gln Phe Asp Asn  
 245 250 255  
 AGA TAC TAC TCC AAT CTT CTG CAG CTC AAT GGC TTA CTT CAG AGT GAC 816  
 10 Arg Tyr Tyr Ser Asn Leu Leu Gln Leu Asn Gly Leu Leu Gln Ser Asp  
 260 265 270  
 CAA GAA CTT TTC TCC ACT CCT GGT GCT GAT ACC ATT CCC ATT GTC AAT 864  
 Gln Glu Leu Phe Ser Thr Pro Gly Ala Asp Thr Ile Pro Ile Val Asn  
 275 280 285  
 15 AGC TTC AGC AGT AAC CAG AAT ACT TTC TTT TCC AAC TTT AGA GTT TCA 912  
 Ser Phe Ser Ser Asn Gln Asn Thr Phe Phe Ser Asn Phe Arg Val Ser  
 290 295 300  
 ATG ATA AAA ATG GGT AAT ATT GGA GTG CTG ACT GGG GAT GAA GGA GAA 960  
 Met Ile Lys Met Gly Asn Ile Gly Val Leu Thr Gly Asp Glu Gly Glu  
 20 305 310 315 320  
 ATT CGC TTG CAA TGT AAT TTT GTG AAT GGA GAC TCG TTT GGA TTA GCT 1008  
 Ile Arg Leu Gln Cys Asn Phe Val Asn Gly Asp Ser Phe Gly Leu Ala  
 325 330 335  
 AGT GTG GCG TCC AAA GAT GCT AAA CAA AAG CTT GTT GCT CAA TCT AAA 1056  
 25 Ser Val Ala Ser Lys Asp Ala Lys Gln Lys Leu Val Ala Gln Ser Lys  
 340 345 350  
 TAAACCAATA ATTAATGGGG ATGTGCATGC TAGCTAGCAT GTAAAGGCAA ATTAGGTTGT 1116  
 AAACCTCTTT GCTAGCTATA TTGAAATAAA CCRAAGGAGT AGTGTGCATG TCAATTGAT 1176  
 TTGCCATGT ACCTCTTGGA ATATTATGTA ATAATTATT GAATCTCTTT AAGGTACTTA 1236  
 30 ATTAATCA 1244

(2) INFORMATION FOR SEQ ID NO: 2:

(1) SEQUENCE CHARACTERISTICS:

35

(A) LENGTH: 3359 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)  
(iii) HYPOTHETICAL: NO  
(iv) ANTI-SENSE: NO  
(ix) FEATURE:  
5       (A) NAME/KEY: exon  
       (B) LOCATION:192..411  
(ix) FEATURE:  
       (A) NAME/KEY: exon  
       (B) LOCATION:1042..1233  
10       (ix) FEATURE:  
          (A) NAME/KEY: exon  
          (B) LOCATION:2264..2429  
(ix) FEATURE:  
          (A) NAME/KEY: exon  
15       (B) LOCATION:2692..3174  
(ix) FEATURE:  
          (A) NAME/KEY: intron  
          (B) LOCATION:412..1042  
(ix) FEATURE:  
20       (A) NAME/KEY: intron  
          (B) LOCATION:1234..2263  
(ix) FEATURE:  
          (A) NAME/KEY: intron  
          (B) LOCATION:2430..2692  
25       (ix) FEATURE:  
          (A) NAME/KEY: CDS  
          (B) LOCATION:192..411  
(ix) FEATURE:  
          (A) NAME/KEY: CDS  
30       (B) LOCATION:1042..1233  
(ix) FEATURE:  
          (A) NAME/KEY: CDS  
          (B) LOCATION:2264..2429  
(ix) FEATURE:  
35       (A) NAME/KEY: CDS  
          (B) LOCATION:2692..3174  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

- 41 -

	GCATCATATC ATAAACAATA CGTACGTGAT ATTATCTAGT GTCTCTCAGT TTA	60
	CTTTTATG AGAAATTATT TTTCTTTAAA AAAAGTTAAT TAATAAAAAC ATTTGCGATA CCGTGAGTTA	120
	CAAGAAATCC GCCGAATTCA TCTCTATAAA TAAAAGGATC TATATGAGAG GTAAAATCAT	180
	ATTA	230
5	Met Gly Ser Met Arg Leu Leu Val Val Ala Leu Leu Cys	
	1 5 10	
	GCA TTT GCT ATG CAT GCA GGT TTT TCA GTC TCT TAT GCT CAG CTT ACT	278
	Ala Phe Ala Met His Ala Gly Phe Ser Val Ser Tyr Ala Gln Leu Thr	
	15 20 25	
10	CCT ACG TTC TAC AGA GAA ACA TGT CCA AAT CTG TTC CCT ATT GTG TTT	326
	Pro Thr Phe Tyr Arg Glu Thr Cys Pro Asn Leu Phe Pro Ile Val Phe	
	30 35 40 45	
	GGA GTA ATC TTC GAT GCT TCT TTC ACC GAT CCC CGA ATC GGG GCC AGT	374
	Gly Val Ile Phe Asp Ala Ser Phe Thr Asp Pro Arg Ile Gly Ala Ser	
15	50 55 60	
	CTC ATG AGG CTT CAT TTT CAT GAT TGC TTT GTT CAA G TACGTACTTT	421
	Leu Met Arg Leu His Phe His Asp Cys Phe Val Gln	
	65 70	
	TTTTTTTCCT TCCAAAATGC CCTGCATATT TAACAAGATT GCTTTGTTCA CCTAGAAAAA	481
20	TGTTGTTTTT TCAACGATCT TACGTACGTT TGTGTTGTTT GAAAAATAAA TCAGAAAGAG	541
	ATCAAGAAAA TAGCTAGAAA GAAAGCAACG TTTTTTTAAA AGGTATTTAG TGTGAGAAAA	601
	ATATTAAAAC TGAAGAGAAA GAAATTAAAT AAGCTTTTCT TGAATGATAT TTACATGTCT	661
	TATTA	721
	CTTA AAGTCACCTT TTTTCTTTAA GTTGTGCTTG AAGAAAAAAG ATGTCTTTCA	781
25	GTTTAGTTTT GATTAATGCT AATTATATTT TTAATTAATT AATTAATACT ATATATCTAT	841
	TTACCATATT AATTATTACT ATATTTTCATG ATGACAACAG ACAAGTATTC TAAAGAGGTA	901
	TCGGTAGATG ATTAATTTTT TTATAAAAAA ATCTTTTGCG TGTATAGATA TTCTTTTATA	961
	ATTGGTGCAG AA	1021
	ACTTGTA TGTCTAATTGC AATTAATCTT ACATTGATTA ACTAATAGCT	1071
30	ATAATCAATA TTTAGGTTAG GTATAGGAGA CAAATCAAGT GATCTGAACA AATTAAGTTG	
	TTATATTGCG ATTGTGACAG GGT TGT GAT GGA TCA GTT TTG CTG AAC AAC	
	Gly Cys Asp Gly Ser Val Leu Leu Asn Asn	
	1 5 10	
	ACT GAT ACA ATA GAA AGC GAG CAA GAT GCA CTT CCA AAT ATC AAC TCA	1119
	Thr Asp Thr Ile Glu Ser Glu Gln Asp Ala Leu Pro Asn Ile Asn Ser	
	15 20 25	
35	ATA AGA GGA TTG GAC GTT GTC AAT GAC ATC AAG ACA GCG GTG GAA AAT	1167
	Ile Arg Gly Leu Asp Val Val Asn Asp Ile Lys Thr Ala Val Glu Asn	
	30 35 40	

- 42 -

	AGT TGT CCA GAC ACA GTT TCT TGT GCT GAT ATT CTT GCT ATT GCA GCT	1215
	Ser Cys Pro Asp Thr Val Ser Cys Ala Asp Ile Leu Ala Ile Ala Ala	
	45 50 55	
	GAA ATA GCT TCT GTT CTG GTAATTAATA ACTCCTAATT AATTCCCAAC	1263
5	Glu Ile Ala Ser Val Leu	
	60	
	CATTAAAAAG TTGCATGATT GGATTCAAAA TTCTATGGTA TTGGGGTTCT GATATAAATT	1323
	TGTAATTAAA TTGCACTAAA AAAAATTATC ATATACTTTT AATAAAAAAA ATTTATCTAA	1383
	TTTAATTTAT TATTAAACT ATTTTAAAA TTCAATCCTA ACTCTTTTTT AATCGGAGCA	1443
10	TGTAAGCTGG CACCCACCGT ATATCGTTGG AAGATGCTAT AAAACCATTT AATTAATGGA	1503
	TGGAATCAGT CAAAACATTT AATTCAAAAT ACTCTTAATT GTGATTAGTA ATCATGTTCTG	1563
	GGCAAGTTAC GTTGTGTATA ATTAATTTGA CTTAATCAGA TAAAAAACA AATGGACGCA	1623
	AGCCGGTTGG TATAGATATC ACTGGCCTGT AGAATATGTG GTTTTTCACG TTTAAATAAA	1683
	AGCTAGCTAC TATATTATAT TTAGTCTTTT TTTTCTTAA ACCCATTTAA CGTGATTTAT	1743
15	TGACTGTGAA ACATGTTTCC ACACACAGGC TTAGAACTC CTCGCAACTA ACATCTCCAA	1803
	AATTTGACTA TTTATTTATG AAGATAATTC ATCTATGATG TTCAACTCTA TTATATATAT	1863
	GTATCATCGC AGTATTAAGA ATTATAATAG TCAAATATAG AAGTATATCG GGTAAATGTA	1923
	GTGCAATGTG CGACCTGTTT CGTGTAATAAT GCTTATTCTA TATAGCTTTT TTTATTGGAA	1983
	AATAACGATG AACTAAAAAC GAAAGGGTAT CATATAGTTT GACTTTTATG TTAGAGAGAG	2043
20	ACATCTTAAT TTGGTCATAT GTTAAATAAT TAATTACAAT GCATACACAA ATATTTATGC	2103
	CATATCTAAA AAATGATAAA ATATCATAGG TATACTCAAC TATATGATAT CCCATAACA	2163
	GAAATTGTAC TTTTCTTCAG GCAATGAACT TAACATTCTT GTTTGCTAAA AACAAACATC	2223
	CACTTAAAGT GGTTCAACAT ATTTATGTAA TAATTTACAG GGA GGA GGT CCA GGA	2278
	Gly Gly Gly Pro Gly	
25	1 5	
	TGG CCA GTT CCA TTA GGA AGA AGG GAC AGC TTA ACA GCA AAC CGA ACC	2326
	Trp Pro Val Pro Leu Gly Arg Arg Asp Ser Leu Thr Ala Asn Arg Thr	
	10 15 20	
	CTT GCA AAT CAA AAC CTT CCA GCA CCT TTC TTC AAC CTC ACT CAA CTT	2374
30	Leu Ala Asn Gln Asn Leu Pro Ala Pro Phe Phe Asn Leu Thr Gln Leu	
	25 30 35	
	AAA GCT TCC TTT GCT GTT CAA GGT CTC AAC ACC CTT GAT TTA GTT ACA	2422
	Lys Ala Ser Phe Ala Val Gln Gly Leu Asn Thr Leu Asp Leu Val Thr	
	40 45 50	
35	CTC TCA G GTATACATAA TCAATTTTTT ATTTGCTATT AGCTAGCAAT AAAAAGTCTC	2479
	Leu Ser	
	55	

	TGATACAGAC ATATTTAGAT AAATTAATTT CTCCATAAAC ATTTATAATA AAATTATCAA	2539
	TTTATGTACT TAAAAATTAT GGATTGAAGC TCTTTTCATC CAACTTTTAC TAAAGTTAAG	2599
	GTGCATATAA TATAAAATAA ACTATCTCTT GTTTCITTATA AAAAGATTGA AGATAAGTTA	2659
	AAGTCTACTT ATAAATCATT AATATATGTA TA GGT GGT CAT ACG TTT GGA AGA	2712
5	Gly Gly His Thr Phe Gly Arg	
	1 5	
	GCT CGG TGC AGT ACA TTC ATA AAC CGA TTA TAC AAC TTC AGC AAC ACT	2760
	Ala Arg Cys Ser Thr Phe Ile Asn Arg Leu Tyr Asn Phe Ser Asn Thr	
	10 15 20	
10	GGA AAC CCT GAT CCA ACT CTG AAC ACA ACA TAC TTA GAA GTA TTG CGT	2808
	Gly Asn Pro Asp Pro Thr Leu Asn Thr Thr Tyr Leu Glu Val Leu Arg	
	25 30 35	
	GCA AGA TGC CCC CAG AAT GCA ACT GGG GAT AAC CTC ACC AAT TTG GAC	2856
	Ala Arg Cys Pro Gln Asn Ala Thr Gly Asp Asn Leu Thr Asn Leu Asp	
15	40 45 50 55	
	CTG AGC ACA CCT GAT CAA TTT GAC AAC AGA TAC TAC TCC AAT CTT CTG	2904
	Leu Ser Thr Pro Asp Gln Phe Asp Asn Arg Tyr Tyr Ser Asn Leu Leu	
	60 65 70	
	CAG CTC AAT GGC TTA CTT CAG AGT GAC CAA GAA CTT TTC TCC ACT CCT	2952
20	Gln Leu Asn Gly Leu Leu Gln Ser Asp Gln Glu Leu Phe Ser Thr Pro	
	75 80 85	
	GGT GCT GAT ACC ATT CCC ATT GTC AAT AGC TTC AGC AGT AAC CAG AAT	3000
	Gly Ala Asp Thr Ile Pro Ile Val Asn Ser Phe Ser Ser Asn Gln Asn	
	90 95 100	
25	ACT TTC TTT TCC AAC TTT AGA GTT TCA ATG ATA AAA ATG GGT AAT ATT	3048
	Thr Phe Phe Ser Asn Phe Arg Val Ser Met Ile Lys Met Gly Asn Ile	
	105 110 115	
	GGA GTG CTG ACT GGG GAT GAA GGA GAA ATT CGC TTG CAA TGT AAT TTT	3096
	Gly Val Leu Thr Gly Asp Glu Gly Glu Ile Arg Leu Gln Cys Asn Phe	
30	120 125 130 135	
	GTG AAT GGA GAC TCG TTT GGA TTA GCT AGT GTG GCG TCC AAA GAT GCT	3144
	Val Asn Gly Asp Ser Phe Gly Leu Ala Ser Val Ala Ser Lys Asp Ala	
	140 145 150	
	AAA CAA AAG CTT GTT GCT CAA TCT AAA TAA ACCAATAATT AATGGGGATG	3194
35	Lys Gln Lys Leu Val Ala Gln Ser Lys *	
	155 160	
	TGCATGCTAG CTAGCATGTA AAGGCAAATT AGGTTGTAAA CCTCTTTGCT AGCTATATTG	3254

2186833

- 44 -

AAATAACCA AAGGAGTAGT GTGCATGTCA ATTCGATTTT GCCATGTACC TCTTGAATA 3314

TTATGTAATA ATTATTGAA TCTCTTTAAG GTACTTAATT AATCA 3359

2223

**THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE  
PROPERTY OF PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

1. An isolated DNA molecule having the nucleotide sequence of SEQ ID NO:1.
2. An isolated DNA molecule comprising a nucleotide sequence substantially homologous to that of SEQ ID NO:2.
3. The isolated DNA molecule of claim 2 having the nucleotide sequence of SEQ ID NO:2.
4. An isolated DNA molecule encoding a DNA regulatory element comprising a nucleotide sequence substantially homologous to that of 1-191 of SEQ ID NO:2.
5. The isolated DNA molecule of claim 4, wherein the DNA regulatory element comprises the nucleotide sequence of 1-191 of SEQ ID NO:2.
6. An isolated DNA molecule of claim 2 comprising the nucleotide sequence of 412-1041 of SEQ ID NO:2.
7. An isolated DNA molecule of claim 2 comprising the nucleotide sequence of 1234-2263 of SEQ ID NO:2.
8. An isolated DNA molecule of claim 2 comprising the nucleotide sequence of 2430-2691 of SEQ ID NO:2.
9. A vector which comprises a DNA molecule selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and nucleotides 1-191 of SEQ ID NO:2.

10. A vector of claim 9 wherein the DNA molecule comprises nucleotides 1-191 of SEQ ID NO:2.
11. A vector of claim 10 which comprises a gene of interest under the control of the DNA molecule.
12. A host cell capable of expressing the DNA molecule within the vector of claim 9.
13. A transgenic plant comprising the vector of claim 9.
14. A method for the production of soybean seed coat peroxidase in a host cell comprising:
  - i) transforming the host cell with the vector comprising an isolated DNA molecule selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2 and nucleotides 1-191 of SEQ ID NO:2, and;
  - ii) culturing the host cell under conditions to allow expression of the soybean seed coat peroxidase.
15. A process for producing a heterologous gene of interest within seed coat cells comprising propagating a transformed plant with the vector of claim 11.



2186833

FIGURE 1

ATGGGTTCCATGCGTCTATT	20
<u>M G S M R L L</u>	
----- prx9+ ----->	
AGTAGTGGCATTGTTGTGTGCATTGCTATGCAGGTTTTTCAGTCTCTTATGCTCA	80
<u>V V A L L C A F A M H A G F S V S Y A Q</u>	1
signal sequence	
GCTTACTCCTACGTTCTACAGAGAAACATGTCCAAATCTGTTCCCTATTGTGTTTGGAGT	140
<u>L T P T F Y R E T C P N L F P I V F G V</u>	21
----- prx12+ ----->	
AATCTTCGATGCTTCTTTACCGATCCCCGAATCGGGGCCAGTCTCATGAGGCTTCATT	200
<u>I F D A S F T D P R I G A S L M R L H F</u>	41
active site	
I	
TCATGATTGCTTTGTTCAAG GTTGTGATGGATCAGTTTGTCTGAACAACACTGATACAAT	260
<u>H D C F V Q G C D G S V L L N N T D T I</u>	61
----- prx10- ----	
AGAAAGCGAGCAAGATGCACCTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAA	320
<u>E S E Q D A L P N I N S I R G L D V V N</u>	81
TGACATCAAGACAGCGGTGGAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATATTCT	380
<u>D I K T A V E N S C P D T V S C A D I L</u>	101
II	
TGCTATTGCAGCTGAAATAGCTTCTGTTCTG GGAGGAGGTCCAGGATGGCCAGTTCCATT	440
<u>A I A A E I A S V L G G G P G W P V P L</u>	121
AGGAAGAAGGGACAGCTTAACAGCAAACCGAACCCTTGCAAATCAAAACCTTCCAGCACC	500
<u>G R R D S L T A N R T L A N Q N L P A P</u>	141
TTTCTTCAACCTCACTCAACTTAAAGCTTCTTTGTGTTCAAGGTCTCAACACCCCTTGA	560
<u>F F N L T Q L K A S F A V Q G L N T L D</u>	161
III	
TTTAGTTACACTCTCAG GTGGTCATACGTTTGGAAAGAGCTCGGTGCAGTACATTTCATAAA	620
<u>L V T L S G G H T F G R A R C S T F I N</u>	181
heme-binding domain	
CCGATTATACAACCTTCAGCAACACTGGAAACCCTGATCCAACCTCTGAACACAACATACTT	680
<u>R L Y N F S N T G N P D P T L N T T Y L</u>	201
AGAAGTATTGCGTGCAAGATGCCCCAGAATGCAACTGGGGATAACCTCACCAATTTGGA	740
<u>E V L R A R C P Q N A T G D N L T N L D</u>	221
CCTGAGCACACCTGATCAATTTGACAACAGATACTACTCCAATCTTCTGCAGCTCAATGG	800
<u>L S T P D Q F D N R Y Y S N L L Q L N G</u>	241
CTTACTTCAGAGTGACCAAGAACTTTTCTCCACTCCTGGTGCTGATACCATTCCCATTGT	860
<u>L L Q S D Q E L F S T P G A D T I P I V</u>	261
<----- prx6- -----	
CAATAGCTTCAGCAGTAACCAGAATACTTTCTTTTCCAACCTTTAGAGTTTCAATGATAAA	920
<u>N S F S S N Q N T F F S N F R V S M I K</u>	281
AATGGGTAATATTGGAGTGTGACTGGGGATGAAGGAGAAATTCGCTTGCAATGTAATT	980
<u>M G N I G V L T G D E G E I R L Q C N F</u>	301
TGTGAATGGAGACTCGTTTGGATTAGCTAGTGTGGCGTCCAAAGATGCTAAACAAAAGCT	1040
<u>V N G D S F G L A S V A S K D A K Q K L</u>	321
TGTTGCTCAATCTAAATAAACCAATAATTAATGGGGATGTGCATGCTAGCTAGCATGTAA	1100
<u>V A Q S K *</u>	326

2186833

FIGURE 1

AGGCAAATTAGGTTGTAAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTAGTG 1160  
TGCATGTCAATTCGATTTTGCCATGTACCTCTTGGAATATTATGTAATAATTATTTGAAT 1220  
CTCTTTAAGGTACTTAATTAATC (A)<sub>n</sub>

FIGURE 2A

L78163	-----ATGGGTTCCATGCGT-CTATTAGTAGTGGCATTGTTG	36
U41657		0
X90693	G---GCAAA-CAATGAACTCCCTTCGTGCTGTAGCAATAG-CTTTGTGC	44
X90694	GCTCTTCAAAACAATGAACTCC-----TTAGCAACTT-CTATGTGG	40
L36156	-----CTCC-----TTAGCAACTT-CTATGTGG	22
X90692	-----AATGCTTGGT-----CTAAGTGCAACAGCTTTTTTGTGTATGG	38
L78163	TGT-----GCATTT-GCTATGCATGCAGGTTTTTCAGT---CTCTTATGC	77
U41657		0
X90693	TGTATTGTG-----GTTGTGCTTGGAGGGTTACCCCTTCTCTTCAAATGC	88
X90694	TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCTTTTCCTCAGATGC	90
L36156	TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCTTTTCCTCAGATGC	72
X90692	TGT-TTGTGCTAAT-----TGGAGGAGTACCCTTTT---CAAATGC	75
L78163	TCAGCTTACTCCTACGTTCTACAGAGAAACATGTCCAAATCTGTTCCCTA	127
U41657		0
X90693	GCAACTTGATCCATCCTTTTACAGGAACACTTGTCCAAATGTTAGTTCCA	138
X90694	ACAACCTAGTCCCACTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA	140
L36156	ACAACCTAGTCCCACTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA	122
X90692	ACAACCTAGATCCTTCATTTTACAACAGTACATGTTCTAATCTTGATTCAA	125
L78163	TTGTGTTTGGAGTAATCTTCGATGCTTCTTTACCCGATCCCCGAATCGGG	177
U41657		0
X90693	TTGTTGCGTGAAGTCATAAGGAGTGTTCCTAAGAAAGATCCTCGTATGCTT	188
X90694	TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT	190
L36156	TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT	172
X90692	TCGTACGTGGTGTGCTCACAATGTTTCAATCTGATCCCAGAATGCTT	175
L78163	GCCAGTCTCATGAGGCTTCATTTTCATGATTGCTTTGTTCAAGGTTGTGA	227
U41657	-----TTTCATGATTGCTTTGTTCAAGGTTGTGA	29
X90693	GCTAGTCTTGTGAGGCTTCACCTTCATGACTGTTTGTTCAGGTTGTGA	238
X90694	GCTAGTCTCGTCAGGCTTCACCTTCATGACTGTTTGTTCAGGTTGTGA	240
L36156	GCTAGTCTCGTCAGGCTTCACCTTCATGACTGTTTGTTCAGGTTGTGA	222
X90692	GGTAGTCTCATCAGGCTACATTTTCATGACTGTTTGTTCAGGTTGCGA	225
L78163	TGGATCAGTTTTGCTGAACAACACTGATACAATAGAAAGCGAGCAAGATG	277
U41657	TGGATCAGTTTTACTGAACAACACTGATACAATAGAAAGCGAGCAAGATG	79
X90693	TGCATCAGTTTTACTAAACAAACACTGATACCGTTGTGAGTGAACAAGATG	288
X90694	TGCCTCAGTTTTGCTGAACAATACTGCTACAATCGTAAGCGAACAACAAG	290
L36156	TGCCTCAGTTTTGCTGAACAATACTGCTACAATCGTAAGCGAACAACAAG	272
X90692	TGCCTCGATTTTGTGCTGAACGATACGGCTACAATAGTGAGCGAGCAAGTG	275
L78163	CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC	327
U41657	CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC	129
X90693	CTTTTCCAAACAGAACTCATTAAAGAGGTTTGGATGTTGTGAATCAAATC	338
X90694	CTTTTCCAAATAACAACCTCTCTAAGAGGTTTGGATGTTGTGAATCAGATC	340
L36156	CTTTTCCAAATAACAACCTCTCTAAGAGGTTTGGATGTTGTGAATCAGATC	322
X90692	CACCACCAATAACAACCTCCATAAGAGGTTTGGATGTTGATAAACCAGATC	325
L78163	AAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT	377
U41657	AAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT	179
X90693	AAAACAGCTGTGGAAAAGGCTTGTCTTAACACAGTTTCTTGTGCTGATAT	388
X90694	AAACTGGCTGTAGAAAGTGCCTTGTCTTAACACAGTTTCTTGTGCTGATAT	390
L36156	AAACTGCTGTAGAAAGTGCCTTGTCTTAACACAGTTTCTTGTGCTGATAT	372
X90692	AAAACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT	375
L78163	TCTTGCTATTGCAGCTGAAATAGCTTCTGTT-CTGGGAGGAGGTCCAGGA	426
U41657	TCTTGCTATTGCAGCTGAAATAGCTTCTGTTGCTGGGAGGAGGTCCAGGA	228
X90693	TCTTGCTCTTTCTGCTGAATTATCATCTACA-CTGGCAGATGGTCCCTGAC	437
X90694	TCTTGCACTTGCT--CAAGCATCCTCTGTT-CTGGCACAAGGTCCCTAGT	439
L36156	TCTTGCACTTGCT--CAAGCATCCTCTGTT-CTGGCACAAGGTCCCTAGT	418
X90692	TCTTGCTCTTTCTGCTGAAATATCATCTGAT-CTGGCAAATGGTCTACT	424

2186833

FIGURE 2A

L78163	-----ATGGGTTCCATGCGT-CTATTAGTAGTGGCATTGTTG	36
U41657	-----	0
X90693	G---GCAAA-CAATGAACCTCCCTTCGTGCTGTAGCAATAG-CTTTGTGTC	44
X90694	GCTCTTCAAACAATGAACCTCC-----TTAGCAACTT-CTATGTGG	40
L36156	-----CTCC-----TTAGCAACTT-CTATGTGG	22
X90692	-----AATGCTTGGT-----CTAAGTGCAACAGCTTTTGTGCTGATGG	38
L78163	TGT-----GCATTT-GCTATGCATGCAGGTTTTTTCAGT---CTCTTATGC	77
U41657	-----	0
X90693	TGTATTGTG-----GTTGTGCTTGGAGGGTTACCCCTTCTCTTCAAATGC	88
X90694	TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCCTTTTCTCAGATGC	90
L36156	TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCCTTTTCTCAGATGC	72
X90692	TGT-TTGTGCTAAT-----TGGAGGAGTACCCCTTTT---CAAATGC	75
L78163	TCAGCTTACTCCTACGTTCTACAGAGAAACATGTCCAAATCTGTTCCTTA	127
U41657	-----	0
X90693	GCAACTTGATCCATCCTTTTACAGGAACACTTGTCCAAATGTTAGTTCCA	138
X90694	ACAACCTAGTCCCACTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA	140
L36156	ACAACCTAGTCCCACTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA	122
X90692	ACAACCTAGATCCTTCATTTTACAACAGTACATGTTCTAATCTTGATTCAA	125
L78163	TTGTGTTTGGAGTAATCTTCGATGCTTCTTTCACCGATCCCCGAATCGGG	177
U41657	-----	0
X90693	TTGTTCTGTAAGTCATAAGGAGTGTCTTAAGAAAGATCCTCGTATGCTT	188
X90694	TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT	190
L36156	TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT	172
X90692	TCGTACGTGGTGTGCTCACAATGTTTCAATCTGATCCCAGAATGCTT	175
L78163	GCCAGTCTCATGAGGCTTCATTTTCATGATTGCTTTGTTCAAGGTTGTGA	227
U41657	-----TTTCATGATTGCTTTGTTCAAGGTTGTGA	29
X90693	GCTAGTCTTGTGTCAGGCTTCATTTTCATGACTGTTTTGTTCAAGGTTGTGA	238
X90694	GCTAGTCTCGTCAAGCTTCATTTTCATGACTGTTTTGTTCTGGGATGTGA	240
L36156	GCTAGTCTCGTCAAGCTTCATTTTCATGACTGTTTTGTTCTGGGATGTGA	222
X90692	GGTAGTCTCATCAGGCTACATTTTCATGACTGTTTTGTTCAAGGTTGCGA	225
L78163	TGGATCAGTTTTGCTGAACAACACTGATACATAAGAAAGCGAGCAAGATG	277
U41657	TGGATCAGTTTTACTGAACAACACTGATACATAAGAAAGCGAGCAAGATG	79
X90693	TGCATCAGTTTTACTAAACAACACTGATACCGTTGTGAGTGAACAAGATG	288
X90694	TGCCTCAGTTTTGCTGAACAATACTGCTACAACTCGTAAGCGAACAACAAG	290
L36156	TGCCTCAGTTTTGCTGAACAATACTGCTACAACTCGTAAGCGAACAACAAG	272
X90692	TGCCTCGATTTTCTGAACGATACGGCTACAACTAGTGAGCGAGCAAGTG	275
L78163	CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC	327
U41657	CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC	129
X90693	CTTTTCCAAACAGAACTCATTAAAGAGGTTTGGATGTTGTGAATCAAATC	338
X90694	CTTTTCCAAATAACAACCTCTCTAAGAGGTTTGGATGTTGTGAATCAGATC	340
L36156	CTTTTCCAAATAACAACCTCTCTAAGGGGTTTGGATGTTGTGAATCAGATC	322
X90692	CACCACCAATAACAACCTCCTAAGAGGTTTGGATGTTGATAAACAGATC	325
L78163	AAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT	377
U41657	AAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT	179
X90693	AAAACAGCTGTGGAAAAGGCTTGTCTTAACACAGTTTCTTGTGCTGATAT	388
X90694	AAACTGGCTGTAGAAAGTGCTTGTCTTAACACAGTTTCTTGTGCTGATAT	390
L36156	AAACTGCTGTAGAAAGTGCTTGTCTTAACACAGTTTCTTGTGCTGATAT	372
X90692	AAAACAGCGGTGGAAAATAGTTGTCTTAACACAGTTTCTTGTGCTGATAT	375
L78163	TCTTGCTATTGCAGCTGAAATAGCTTCTGTT-CTGGGAGGAGGTCCAGGA	426
U41657	TCTTGCTATTGCAGCTGAAATAGCTTCTGTTGCTGGGAGGAGGTC-AGGA	228
X90693	TCTTGCTCTTTCTGCTGAATTATCATCTACA-CTGGCAGATGGTCCTGAC	437
X90694	TCTTGCACTTGTGCTCAAGCATCCTCTGTT-CTGGCACAAGGTCCTAGT	439
L36156	TCTTGCACTTGTGCT---CAAGCATCCTCTGTT-CTGGCACAAGGTCCTAGT	418
X90692	TCTTGCTCTTTCTGCTGAAATATCATCTGAT-CTGGCAAATGGTCCTACT	424

FIGURE 2A

L78163	CTTTTCTCCACTCCTGGTGTGATACCATTCCCATTGTCAATAGCTTCAG	872
U41657	CGTTTCTCCACTCCTGGTGTGATACCATTCC-ATTGTCAATAGCTTCAG	667
X90693	TTGTTCTCAACATCTGGTTCAGATACCATTAGCATTGTCAACAAATTCGC	883
X90694	TTGTTCTCAACTTCTGGTGCAGATACCATTAGCATTGTCAACAAATTCAG	885
L36156	TTGTTCTCAACTTCTGGTGCAGATACCATTAGCATTGTGCGAAATTCAG	864
X90692	CTTTTTCAGAAATGGTCTGACACTATTCTATTGTCAATAGTTTCGC	870
	..** ** * . ***** ** * . ***** * . *	
L78163	CAGTAACCAGAATACTTTCTTTTCCAACTTTAGAGTTTCAATGATAAAAA	922
U41657	CG--AACCAGAATACTTTCTTTTCCAACTTTAGAGTTTCAATGATAAAAA	715
X90693	AACCGATCAAAAAGCTTTTGTGAGAGCTTTAGGGCTGCTATGATCAAAA	933
X90694	CACCGATCAAAAAGCTTTTGTGAGAGCTTTAAGGCTGCAATGATTAAAA	935
L36156	CACCGATCAAAAAGCTTTTGTGAGAGCTTTAAGGCTGCAATGATTAAAA	914
X90692	CAATAATCAAACTCTCTTTTGAATAATTTGTAGCCTCAATGATAAAAA	920
	. * ** * . ** * . * . * . * . * . * . * . *	
L78163	TGGGTAATATTGGAGTGTGACTGGGGATGAAGGAGAAATTCGCTTGCAA	972
U41657	TGGGTAATATTGGAGTGTGACTGGGGATGAAGGAGAAATTCGCTTGCAA	765
X90693	TGGGAAATATTGGTGTGTAAACCGGAACCAAGGAGAGATTAGAAAACAA	983
X90694	TGGGCAATATTGGTGTGTAAACAGGGACAAAAGGAGAGATTAGAAAACAA	985
L36156	TGGGCAATATTGGTGTGTAAACAGGGACAAAAGGAGAGATTAGAAAACAA	964
X90692	TGGGTAATATTGGAGTTTAACTGGATCTCAAGGTGAATTAGAACACAG	970
	**** ***** ** * . * . * . * . * . * . *	
L78163	TGTAATTTTGTGAA---TGGAGACTCGT-----TTGGATTAGC	1007
U41657	TGTAATTTTGTGAA---TGGAGACTCGT-----TTGGATTAGC	800
X90693	TGCAACTTTGTTAATT-----CAAAATCAGCAGAACTTGGTCTTAT	1024
X90694	TGCAACTTTGTGAACCTTTGTGAACCTCAAATTCGCAGAACTAGATTAGC	1035
L36156	TGCAACTT-----TGTGAACCTCAAATTCGCAGAACTAGATTAGC	1005
X90692	TG-----TAATGCTGTGAATGGGAATTCCTC-----TGGATTGGC	1005
	** .. * . * . *	
L78163	TAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTGCTCAATCTAAAT	1057
U41657	TAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTGCTCAATCTAAAT	850
X90693	CAATGTTGCCTC---AGCAG--ATTCATCTG-AGGAGGGTATGGTTAG--	1066
X90694	CACCATAGCATCCATAGTAG--AATCATTAG-AGGATGGTATTGCTAGTG	1082
L36156	CACCATAGCATCCATAGTAG--AATCATTAG-AGGATGGAATTGCTAGTG	1052
X90692	TACTGTAGTCACCAA--AG--AATCATCAG-AAGATGGAATGGCTAGCT	1049
	* . * . * . * . * . * . * . * . * . *	
L78163	AAACCAATAATTAATGGGGATGTGCATGCTAGCTAGCATGTAAAGGCCAAA	1107
U41657	AAACCAATAATTAATGGGGATGTGCATGCTAGCTAGCATGTAAAGGCCAAA	900
X90693	-----CTCAATGTAAA-TG-TAG	1082
X90694	TAATATAAATAAATTAG-----CGTAAATGCACCTATTGAA-ATCTTG	1124
L36156	TAATATAAATAAATTAG-----CGTAAATGCACCTATTGAA-ATCTTG	1094
X90692	CATTCTAAAT--ATAAG-----CTTGGAATAATTTGAAGAGGTTCTAT	1090
	. . . . * . . . .	
L78163	TTAGGTTGTAAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTA	1157
U41657	TTAGGTTG-AAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTA	949
X90693	T--GATTGGAAGCAACTAA--TAAATTAAGAAGCTATAAC-----T	1119
X90694	T--GACTAGATGCCACTAA--TAAAT--AAGTTATAAC-----T	1157
L36156	T--GACTAGATCCCCTAA--TAAAT--AAGTTATAAC-----T	1127
X90692	A--ATTTGTGCATACATA--TATGGTATGTG-----	1118
	. . . * . . . * . . * . .	
L78163	GTGTGCATGTCAATTCGATTTTGC-CATGTACCTCTTGGAAATAT-----	1200
U41657	GTGTGCATGTCAATTCGATTTTGC-CATGTACCTCTTGGAAATATTATGTA	998
X90693	ATGCACATT-CATGGTATGTGTGAGATAGTTATTAGATGCTTTGTGAGCA	1168
X90694	AGGCACATTTTGTGCACTTGAAATTTTGTGAGCT--GTATATGAG-----	1200
L36156	AGGCACATTTTGTGCACTTGAAATTTTGTGAGCT--GTATATGAG-----	1177
X90692	-----CATGTGGTGTA--TTATGTTTTGTGTATGTTCTCAAGTTGATCA	1161
	** . . . . * . . . . * . . . .	

2186833

FIGURE 2A

L78163	-----	1200
U41657	ATAATTATTTGAATCTC-----AAAAAAAAAAAAAAAA	1031
X90693	AAAATCTTTTGGATTTC---ATTGAAGTGTTTCT---	1200
X90694	-----	1200
L36156	TGT-TCTT-----C-----TTGGTATTATACTA--T	1200
X90692	GGGA-CTGTAGAAGCTCCCTAATAATATTTGTGTCAAAGT	1200

FIGURE 2B

L78163	MGSMRLLVALLCAFAMHAGFSVSY---AQLTPTFYRETCPNLFPIVFGV	47
U41657	-----	0
X90693	MNSLRVAIAIACCIV--VVLGGLPFSSNAQLDPSFYRNTCPNVSSIVREV	48
X90694	MNSL---ATSMWCVVLLVVLGGLPFSSDAQLSPTFYSKTCPTVSSIVSNV	47
L36156	M-----WCVVLLVVLGGLPFSSDAQLSPTFYSKTCPTVSSIVSNV	40
X90692	MLGLSATA---FCCMVFLVIGGVPPS-NAQLDPSFYNSTCSNLDIVRGV	46
L78163	IFDASFTDPRIGASLMRLHFHDCFVQCGDGSVLLNNTDTIESEQDALPNI	97
U41657	-----FHDCFVQCGDGSVLLNNTDTIESEQDALPNI	31
X90693	IRSVSKKDPRLASLVRHLHFHDCFVQCGDASVLLNKTDTVVSEQDAFPNR	98
X90694	LTVNSKTDPRMLASLVRHLHFHDCFVQCGDASVLLNNTATIVSEQQAFPNR	97
L36156	LTVNSKTDPRMLASLVRHLHFHDCFVQCGDASVLLNNTATIVSEQQAFPNR	90
X90692	LTVNSQSDPRMLGSLIRLHFHDCFVQCGDASILLNDTATIVSEQSAPPNR	96
	***** **	
L78163	NSIRGLDVVNDIKTAVENSCPDTVSCADILAIABIASVLGGGPGWVPVL	147
U41657	NSIRGLDVVNDIKTAVENSCPDTVSCADILAIABIASVAGRRSGWVPVL	81
X90693	NSIRGLDVVNDIKTAVEKACPDTVSCADILALSSELSTLADGPDWKVPL	148
X90694	NSIRGLDVVNDIKTAVEVPCPDTVSCADILALAAQASSVLAQGPSWTVPVL	147
L36156	NSIRGLDVVNDIKTAVESACPDTVSCADILALA-QASSVLAQGPSWTVPVL	139
X90692	NSIRGLDVINQIKTAVENACPDTVSCADILALSSELSTLANGPTWQVPL	146
	***** **	
L78163	GRRDSLANTANRTLANQNLPAPFFNLTLQKASFAVQGLNTLDLVTLSGGHTF	197
U41657	GRRDSLANTANRTLANQNLPAPFFNLTLQKASFAVQGLNTLDLVTLSGGHTS	131
X90693	GRRDGLTANRTLANQNLPAPFNTTDLQKAAFAAQGLDITDLVALSGAHTF	198
X90694	GRRDGLTANRTLANQNLPAPFNSLDLQKAAFTAQGLITPVLVALSGAHTF	197
L36156	GRRDGLTANRTLANQNLPAPFNSLDLQKAAFTAQGLITPVLVALSGAHTF	189
X90692	GRRDSLANTANRTLANQNLPAPFNTTDLQKAAFTAQGLITPVLVALSGAHTF	196
	***** **	
L78163	GRARCSTFINRLYNFSNTGNPDPTLNTTYLEVLRARCPQONATGDNLTNLD	247
U41657	GRARCSTFINRLYNFSNTGLIH--LDTTYLEVLRARCPQONATGDNLTNLD	179
X90693	GRAHCSLFVSRLYNFSSTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD	248
X90694	GRAHCAQFVSRLYNFSSTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD	247
L36156	GRAHCAQFVSRLYNFSSTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD	239
X90692	GRGQCRFFVDRLYNFSNTGNPDSTLNTTYLQQLRTICPNGGPGTNLTNFD	246
	***** **	
L78163	LSTPDQFDNRYYSNLLQLNGLLQSDQELFSTPGADTIPIVNSFSSNQNTF	297
U41657	LSTPDQFDNRYYSNLLQLNGLLQSDQERFSTPGADTIPLSIA-SANQNTF	228
X90693	PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGSDTISIVNKFATDQKAF	298
X90694	PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGSDTISIVNKFATDQKAF	297
L36156	PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGSDTISIVNKFATDQKAF	289
X90692	PTTPDFTDSNYYSNLQVKKGLFQSDQELFSRNGSDTISIVNSFANNQTLF	296
	***** **	
L78163	FSNFRVSMIKMGNIGVLTGDEGEIRLQCNFVN-----GDSFGLASVAS-K	341
U41657	FSNFRVSMIKMGNIGVLTGDEGEIRLQCNFVN-----GDSFGLASVAS-K	272
X90693	FESFRAAMIKMGNIGVLTGNQGEIRKQCNFVN---SKSAELGLINVAS-A	344
X90694	FESFKAAMIKMGNIGVLTGKGEIRKQCNFVN---SNSAELDLATIASIV	347
L36156	FESFKAAMIKMGNIGVLTGKGEIRKQCNFVN---SNSAELDLATIASIV	336
X90692	FENFVASMIMKMGNIGVLTGSQGEIRTQCNFVN-----GNSSGLATVVT-K	340
	***** **	
L78163	DAKQKLVAQSK	352
U41657	DAKQKLVAQSK	283
X90693	DSSEEGMVSSM	355
X90694	ESLEDGIASVI	358
L36156	ESLEDGIASVI	347
X90692	ESSEDGMASF	351
	***** **	

FIGURE 3

10 20 30 40 50 60  
 1 GCATCATATCATAACAATACGTACGTGATATTATCTAGTGTCTCTCAGTTTACTTTTATG  
 61 AGAAATTATTTTCTTTAAAAAAGTTAATTAATAAAAAACATTGCGATACCGTGAGTTA  
 121 CAAGAAATCCGCCGAATTCATCTCTATAAATAAAAGGATCTATATGAGAGGTAAATCAT  
 181 ATTAACCTCAAATGGGTCCATGCGTCTATTAGTAGTGGCATTGTTGTGTGCATTTGCTA  
 241 TGCATGCAGGTTTTTTCAGTCTCTTATGCTCAGCTTACTCCTACGTTCTACAGAGAAACAT  
 301 GTCCAAATCTGTTCCCTATTGTGTTTGGAGTAATCTTCGATGCTTCTTTACCGATCCCC  
 361 GAATCGGGGCCAGTCTCATGAGGCTTCATTTTCATGATTGCTTGTTCAGTACGTACTT  
 421 TTTTTTTCCTTCCAAAATGCCCTGCATATTTAACAAGATTGCTTGTTCACCTAGAAAA  
 481 ATGTGTTTTTTTCAACGATCTTACGTACGTTTGTGTTTGAATAAATAAATCAGAAAGA  
 541 GATCAAGAAAATAGCTAGAAAGAAAGCAACGTTTTTTTAAAGGTATTTAGTGTGAGAAA  
 601 AATATTAAACTGAAGAGAAAGAAATTAATAAGCTTTTCTTGAATGATATTTACATGTC  
 661 TTATTAACTTAAAGTCACCTTTTTTCTTTAAGTTGTGCTTGAAGAAAAAGATGCTTTT  
 721 AGTTTATGTTTGAATTAATGCTAATTATATTTTAAATTAATTAATACTATATATCTA  
 781 TTTACCATATTAATTATTACTATATTTTCATGATGACAACAGACAAGTATTTCTAAAGAGGT  
 841 ATCGGTAGATGATTATTTTTTATAAAAAAATCTTTTGCCTGTATAGATATCTTTTAT  
 901 AATTGGTGCAAACTTGAATGCTAATTGCAATTAATCTTACATTGATTAACTAATAGC  
 961 TATAATCAATATTTAGGTTAGGTATAGGAGACAAATCAAGTGATCTGAACAAATTAAGTT  
 1021 GTTATATTTGCATTGTGACAGGGTTGTGATGGATCAGTTTTGCTGAACAACTGATACA  
 1081 ATAGAAAGCGAGCAAGATGCACCTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTC  
 1141 AATGACATCAAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTGTGCTGATATT  
 1201 CTTGCTATTGCAGCTGAAATAGCTTCTGTTCTGGTAATTAATACTCCTAATTAATTTCC  
 1261 AACCATTAAAAAGTTGCATGATTGGATTCAAAATTCTATGGTATTGGGGTTCTGATATAA  
 1321 ATTTGTAATTAAATTCACATAAAAAAATTTATCATATACTTTAAATAAAAAAATTTATC  
 1381 TAATTTAATTTAATTAAACTATTTTTAAAAATTCAATCCTAATCTTTTTTAATCGGA  
 1441 GCATGTAAGCTGGCACCACCGTATATCGTTGGAAGATGCTATAAAACCATTTAATTAAT  
 1501 GGATGGAATCAGTCAAAACATTTAATTCAAAATACTCTTAATTGTGATTAGTAATCATGT  
 1561 TCGGGCAAGTTACGTTGTGTATAATTAATTGACTTAATCAGATAAAAAAACAATGGAC  
 1621 GCAAGCCGTTGGTATAGATATCACTGGCCTGTAGAAATATGTGGTTTTTCACGTTTAAAT  
 1681 AAAAGCTAGCTACTATATTATATTAGTCTTTTTTTTTCTTAAACCATTTAACGTGATT  
 1741 TATTGACTGTGAAACATGTTTCCACACACAGGCTTAGAAACTCCTCGCAACTAACATCTC  
 1801 CAAAATTTGACTATTTATTTATGAAGATAATTCATCTATGATGTTCAACTCTATTATATA  
 1861 TATGTATCATCGCAGTATTAAGAATTATAATAGTCAAATATAGAAGTATATCGGGTAAAT  
 1921 GTAGTTGCATGTGCGACCTGTTTCGTGTAAATGCTTATTCTATATAGCTTTTTTTATTG  
 1981 GAAATAACGATGAACTAAAAACGAAAGGGTATCATATAGTTTGAATTTTATGTTAGAGA  
 2041 GAGACATCTTAATTTGGTCATATGTTAAATAATTAATTAACAATGCATACACAAATATTTA  
 2101 TGCCATATCTAAAAATGATAAAATATCATAGGTATACTCAACTATATGATATCCCCATA  
 2161 ACAGAAATTGTACTTTTTCTTCAGGCAATGAACTTAACATTTCTGTTTGTAAAAACAAAC  
 2221 ATCCACTTAAAGTGGTTCAACATATTTATGTAATAATTTACAGGGAGGAGGTCCAGGATG  
 2281 GCCAGTTCCATTAGGAAGAAGGGACAGCTTAACAGCAAACCGAACCTTGCAATCAAAA  
 2341 CTTCCAGCACCTTTCTTCAACCTCACTCAACTTAAAGCTTCTTTGCTGTTCAAGGTCT  
 2401 CAACACCCTTGATTAGTTTACACTCTCAGGTATACATAATCAATTTTTTATTGCTATTA  
 2461 GCTAGCAATAAAAAGTCTCTGATACAGACATATTTAGATAAATTAATTTCTCCATAAACA  
 2521 TTTATAATAAAATTTATCAATTTATGTAATTTAATAATTTATGGATTGAAGCTCTTTTATCC  
 2581 AACTTTTACTAAAGTTAAGGTGCATATAATATAAAATAAACTATCTCTGTTTCTTATAA  
 2641 AAAGATTGAAGATAAGTTAAAGTCTACTTATAAATCATTAAATATATGATAGGTGGTCAT  
 2701 ACGTTTGAAGAGCTCGGTGCAGTACATTCTATAAACCGATTATACAACCTCAGCAACACT  
 2761 GGAAACCTGATCCAACTCTGAACACAACATACTTAGAAGTATTGCGTGCAAGATGCCCC  
 2821 CAGAATGCAACTGGGATAACCTCACCAATTTGGACCTGAGCACACCTGATCAATTTGAC  
 2881 AACAGATACTACTCCAATCTTCTGCAGCTCAATGGCTTACTTCAGAGTGACCAAGAATTT  
 2941 TTCTCCACTCCTGGTGCTGATACCATTCCCATTGTCAATAGCTTCAGCAGTAACCAGAA  
 3001 ACTTTCTTTTCCAATTTAGAGTTTCAATGATAAAATGGGTAATATTGGAGTGTGACT  
 3061 GGGGATGAAGGAGAAATTCGCTTGCAATGTAATTTGTGAATGGAGACTCGTTTGGATTA  
 3121 GCTAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTGCTCAATCTAAATAAACCAAT  
 3181 AATTAATGGGGATGTGCATGCTAGCTAGCATGTAAGGCAAAATAGGTTGTAAACCTCTT  
 3241 TGCTAGCTATATTGAATAAACCAGGAGTAGTGTGCATGTCAATTCGATTTTGCCATG  
 3301 TACCTCTTGAATATTATGTAATAATTTGAATCTCTTTAAGGTACTTAATTAATCA



FIGURE 4

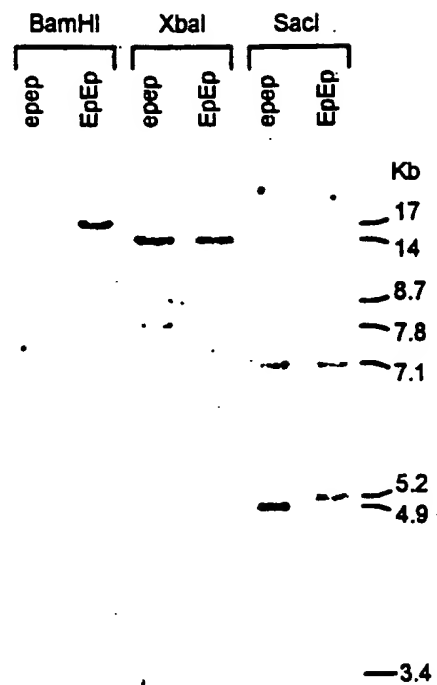
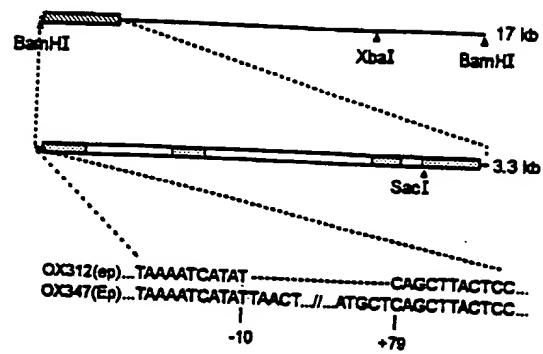
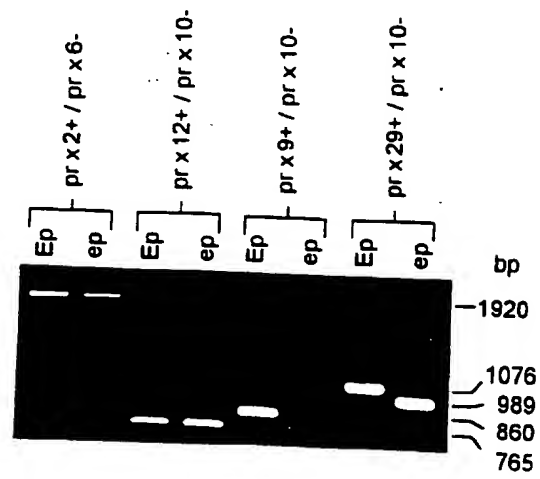


FIGURE 5



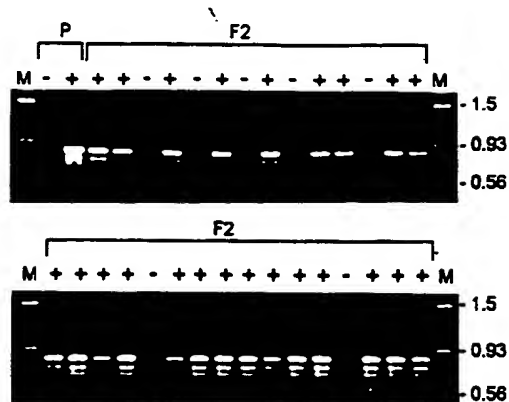
2186833

FIGURE 6



2186833

**FIGURE 7**



2186833

FIGURE 8

